

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2001 年 3 月 8 日 (08.03.2001)

PCT

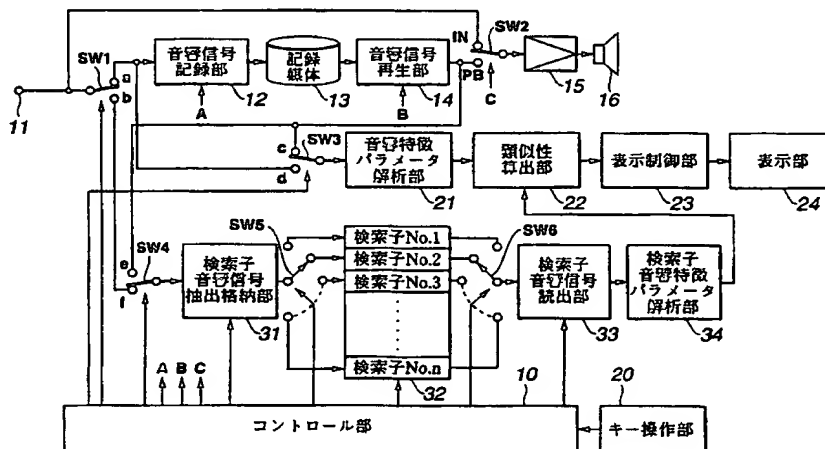
(10) 国際公開番号  
WO 01/16935 A1

- (51) 国際特許分類: G10L 15/00, (71) 出願人 (米国を除く全ての指定国について): ソニー株式会社 (SONY CORPORATION) [JP/JP]; 〒141-0001 東京都品川区北品川6丁目7番35号 Tokyo (JP).
- (21) 国際出願番号: PCT/JP00/05771 (72) 発明者; および
- (22) 国際出願日: 2000 年 8 月 25 日 (25.08.2000) (75) 発明者/出願人 (米国についてのみ): 三浦雅美 (MIURA, Masayoshi) [JP/JP]; 矢部 進 (YABE, Susumu) [JP/JP]; 〒141-0001 東京都品川区北品川6丁目7番35号 ソニー株式会社内 Tokyo (JP).
- (25) 国際出願の言語: 日本語 (74) 代理人: 小池 晃, 外 (KOIKE, Akira et al.); 〒105-0001 東京都港区虎ノ門二丁目6番4号 第11森ビル Tokyo (JP).
- (26) 国際公開の言語: 日本語 (81) 指定国 (国内): CN, KR, US.
- (30) 優先権データ:  
特願平11/239145 1999 年 8 月 26 日 (26.08.1999) JP

[続葉有]

(54) Title: INFORMATION RETRIEVING/PROCESSING METHOD, RETRIEVING/PROCESSING DEVICE, STORING METHOD AND STORING DEVICE

(54) 発明の名称: 情報の検索処理方法、検索処理装置、蓄積方法及び蓄積装置



- 12...ACOUSTIC SIGNAL RECORDING UNIT  
13...RECORDING MEDIUM  
14...ACOUSTIC SIGNAL REPRODUCING UNIT  
21...ACOUSTIC FEATURE PARAMETER ANALYZING UNIT  
22...SIMILARITY COMPUTING UNIT  
23...DISPLAY CONTROL UNIT  
24...DISPLAY UNIT  
31...RETRIEVER ACOUSTIC SIGNAL EXTRACTING/STORING UNIT  
32...RETRIEVER NO.1 RETRIEVER NO.2  
RETRIEVER NO.3 RETRIEVER NO.n  
33...RETRIEVER ACOUSTIC SIGNAL READING UNIT  
34...RETRIEVER ACOUSTIC FEATURE PARAMETER ANALYZING UNIT  
10...CONTROL UNIT  
20...KEY OPERATION UNIT

(57) Abstract: An acoustic feature parameter analyzing unit provided in the front stage of a similarity computing unit analyzes an acoustic signal to be retrieved to obtain an acoustic feature parameter. A retriever acoustic signal reading unit reads the acoustic signal of a retriever from a retriever storing area in a retriever memory for analysis by a retriever acoustic feature parameter analyzing unit to obtain a retriever acoustic feature parameter. The similarity computing unit compares the acoustic feature parameter obtained by the acoustic feature parameter analyzing unit with the retriever acoustic feature parameter from the retriever acoustic feature parameter analyzing unit to compute the similarity between the two and retrieves a portion similar to an acoustic signal from a sound source of the retriever from an acoustic signal to be retrieved.

[続葉有]



(84) 指定国 (広域): ヨーロッパ特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

添付公開書類:

— 国際調査報告書

(57) 要約:

検索対象の音響信号について、類似性算出部の前段に設けた音響特徴パラメータ解析部により解析して音響特徴パラメータを得る。また、検索子音響信号読出部により、検索子メモリの検索子格納エリアから検索子の音響信号を読み出し、この検索子の音響信号について、検索子音響特徴パラメータ解析部により解析して、検索子音響特徴パラメータを得る。そして、類似性算出部では、上記音響特徴パラメータ解析部により得た音響特徴パラメータと、検索子音響特徴パラメータ読出部からの検索子音響特徴パラメータとを比較して、両者の類似性を算出し、検索対象の音響信号中から検索子の音源からの音響信号と類似する部分を検索する。

## 明細書

情報の検索処理方法、検索処理装置、蓄積方法及び蓄積装置

### 技術分野

この発明は、検索対象の音響信号や映像信号から、例えば特定の楽器、特定の人声、動物の鳴き声等の部分や特定の人物画像、特定のシーンなどの部分を検索する検索処理方法及び装置、また、この検索処理方法及び装置を用いた情報蓄積方法及び装置に関する。

### 背景技術

例えば、会議録音を行った場合、記録された音響信号について、記録後に、特定の発言部分などを検索したりする要求がある。

このように、記録された音響信号について、使用者が希望する音響信号部分を検索する場合、従来は、記録時にマーカを打ち込んでおき、再生時に、この打ち込みマーカを検出することにより、希望する音響信号部分を検索するようにしている。

また、記録時に、後で検索したい音響信号部分が判るように、カウンタ数値や、記録開始時点からの時間を、それぞれの音響信号部分と対応を付けて記録するようにする方法も知られている。

しかしながら、上述の方法の場合、記録時に、後で検索するであ

ろう音響信号部分を判断してマーカを打ち込んだり、カウンタ数値や時間を記録したりする必要があり、使い勝手について、次のような問題があった。

1. 記録時に、マーカを使用者が打ち込んだり、カウンタ数値や時間を別途に記録しなければならず、マーカ打ち込み操作やカウンタ数値や時間を記録する操作が厄介である。

2. マーカやカウンタ数値、時間が記録されていない箇所は、検索することはできない。

3. 特定の話者のみ、あるいは特定の楽器のみなどの部分を検索することができない。このため、例えば会議記録において発言区間を話者毎に分類したり、演奏区間を楽器毎に分類したりするなどということができない。

4. 一旦、記録して、マーカやカウンタ数値等を記録する必要があるため、記録を行わずに、複数の検索子について検索して分類するなどということはできなかった。

## 発明の開示

この発明は、以上の点にかんがみ、記録時に、マーカを打ち込んだり、カウンタ数値や時間を記録する操作が必要なく、しかも、特定の話者や楽器などの部分を検索することができる方法及び装置を提供することを目的とする。

また、音響信号を記録しなくても、検索対象の音響信号から、特定の音源などの音響信号を検索子とする検索が行えるようにする方

法及び装置を提供することを目的とする。

本発明に係る音響検索処理方法は、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記検索対象の音響信号中から前記検索子の音響信号と類似する部分を検索することを特徴とする。

また、本発明に係る音響情報蓄積方法は、入力音響信号を記録媒体に記録するとともに、前記入力音響信号を解析して、音源の構造的な要素に依存する音響特徴パラメータを得、検索子とする1～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記検索対象の音響信号中から前記検索子の音響信号と類似する部分を検索し、その検索結果を、前記検索子及び前記入力音響信号の前記記録媒体上の記録位置に関連させて記録するようにすることを特徴とする。

また、本発明に係る音響情報検索装置は、検索対象の音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段の結果を表示する結果表示手段とを備えることを特徴

とする。

また、本発明に係る音響情報検索装置は、検索対象の音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段での検索結果に基づいて、前記類似する部分のみを抽出して再生する再生手段とを備えることを特徴とする。

また、本発明に係る音響情報蓄積装置は、入力音響信号を記録媒体に記録する記録手段と、前記入力音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする1～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段とを備え、前記検索手段の検索結果を、前記検索子及び前記入力音響信号の前記記録媒体上の記録位置に関連させて記録するようにすることを特徴とする。

また、本発明に係る音響映像検索処理方法は、映像信号と、その映像信号に関連する音響信号とからなる入力信号から、所定の部分を検索する音響映像検索処理方法において、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パ

ラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索し、その検索した部分を、前記入力信号についての検索結果とすることを特徴とする。

また、本発明に係る音響映像検索処理方法は、映像信号と、その映像信号に関連する音響信号とからなる入力信号から、所定の部分を検索する音響映像検索処理方法において、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索するとともに、検索子とする1～複数の映像信号についての、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、検索対象の前記入力信号中の映像信号を解析して得た前記映像特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の映像信号中から前記検索子の画像と類似する画像部分を検索し、前記音響信号についての検索結果と、前記映像信号についての検索結果とに基づいて、前記入力信号中から所定の部分を検索することを特徴とする。

また、本発明に係る音響映像情報蓄積方法は、映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録するとともに、前記入力信号中の音響信号を解析して、音源の構

造的な要素に依存する音響特徴パラメータを得、検索子とする 1 ～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力信号中の音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索し、その検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする。

また、本発明に係る音響映像情報蓄積方法は、映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録するとともに、検索子とする 1 ～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索し、さらに、検索子とする 1 ～複数の映像信号についての、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、検索対象の前記入力信号中の映像信号を解析して得た前記映像特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の映像信号中から前記検索子の画像と類似する画像部分を検索し、前記音響信号及び映像信号についての検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする。

また、本発明に係る音響映像情報検索装置は、映像信号と、その



映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段の結果を表示する結果表示手段とを備えることを特徴とする。

また、本発明に係る音響映像情報検索装置は、映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段での検索結果に基づいて、前記類似する部分に対応する前記入力信号部分を抽出して再生する再生手段とを備えることを特徴とする。

また、本発明に係る音響映像情報検索装置は、映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音源についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音

響特徴パラメータ解析手段からの前記入力信号中の音響信号の音響特徴パラメータとを比較して、前記入力信号中の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する音響部分検索手段と、前記入力信号中の映像信号を、画像の特徴を表す映像特徴パラメータに解析する映像特徴パラメータ解析手段と、検索子とする1～複数の映像信号についての前記映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、前記映像特徴パラメータ解析手段からの前記入力信号中の映像信号の映像特徴パラメータとを比較して、前記入力信号中の映像信号中から前記検索子の映像信号と類似する部分を検索する映像部分検索手段と、前記音響部分検索手段及び映像部分検索手段の結果を表示する結果表示手段とを備えることを特徴とする。

また、本発明に係る音響映像情報検索装置は、映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音源についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記入力信号中の音響信号の音響特徴パラメータとを比較して、前記入力信号中の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する音響部分検索手段と、前記入力信号中の映像信号を、画像の特徴を表す映像特徴パラメータに解析する映像特徴パラメータ解析手段と、検索子とする1～複数の映像信号についての前記映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、前記映像特徴パラメータ解析手段からの前記入力信号中の映像信号の映像特徴パラメータ

タとを比較して、前記入力信号中の映像信号中から前記検索子の映像信号と類似する部分を検索する映像部分検索手段と、前記音響部分検索手段及び映像部分検索手段での検索結果に基づいて決定される前記入力信号部分を抽出して再生する再生手段とを備えることを特徴とする。

また、本発明に係る音響映像情報蓄積装置は、映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録する記録手段と、前記入力信号中の音響信号を解析して、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする1～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力信号中の音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索する音響部分検索手段とを備え、前記音響部分検索手段の検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする。

さらに、本発明に係る音響映像情報蓄積装置は、映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録する記録手段と、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索する音響部分検索手段と、

検索子とする 1～複数の映像信号についての、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、検索対象の前記入力信号中の映像信号を解析して得た前記映像特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の映像信号中から前記検索子の画像と類似する画像部分を検索する映像部分検索手段とを備え、前記音響部分検索手段及び映像部分検索手段の検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする。

#### 図面の簡単な説明

図 1 は、この発明による音響検索処理装置の第 1 の実施の形態のブロック図である。

図 2 A 及び図 2 B は、男性音声の音響的特徴を示すための周波数成分分布及び時間波形を示す図である。

図 3 A 及び図 3 B は、女性音声の音響的特徴を示すための周波数成分分布及び時間波形を示す図である。

図 4 A 及び図 4 B は、サキソホンの音響的特徴を示すための周波数成分分布及び時間波形を示す図である。

図 5 A 及び図 5 B は、ベースギターの音響的特徴を示すための周波数成分分布及び時間波形を示す図である。

図 6 は、第 1 の実施の形態における検索子の登録処理動作を説明するためのフローチャートである。

図 7 は、第 1 の実施の形態における検索処理動作を説明するためのフローチャートである。

図 8 A 及び図 8 B は、第 1 の実施の形態における検索処理結果の表示例を示す図である。

図 9 は、第 1 の実施の形態における検索処理結果の表示例の他の例を示す図である。

図 10 は、第 1 の実施の形態における検索処理結果の表示例の他の例を示す図である。

図 11 は、この発明による音響検索処理装置の第 2 の実施の形態のブロック図である。

図 12 は、この発明による音響検索処理装置の第 3 の実施の形態のブロック図である。

図 13 は、この発明による音響検索処理装置の第 4 の実施の形態のブロック図である。

図 14 は、この発明による音響検索処理装置の第 5 の実施の形態の要部の説明に用いる図である。

図 15 は、この発明による音響検索処理装置の第 5 の実施の形態の要部の説明に用いるフローチャートである。

図 16 は、この発明による音響映像検索処理装置の第 1 の実施の形態のブロック図である。

図 17 は、この発明による音響映像検索処理装置の第 1 の実施の形態の説明に用いる図である。

発明を実施するための最良の形態

以下、本発明を実施するための最良の形態について図面を参照しながら詳細に説明する。

以下、この発明の実施の形態を、図を参照しながら説明する。

#### [第1の実施の形態]

図1は、この発明の第1の実施の形態のブロック図である。この第1の実施の形態では、検索対象の入力音響信号、例えばマイクロホンで収音した入力オーディオ信号を記録し、再生する機能を備える。この実施の形態では、再生音響信号も、検索対象の音響信号とする機能を備える。また、検索対象の入力音響信号又は再生音響信号から検索子の音響信号を抽出して記憶蓄積する機能と、この蓄積された検索子の音響信号を用いて入力音響信号又は再生音響信号について検索する機能と、さらに、検索結果を表示部に表示する機能を備える。

入力音響信号は、入力端子11を通じて、スイッチ回路SW1に供給される。このスイッチ回路SW1は、コントロール部10からの切り換え信号により、入力音響信号の記録時及び入力音響信号についての検索時には、端子a側に切り換えられ、入力音響信号から検索子の音響信号を抽出するときには、端子b側に切り換えられる。

そして、スイッチ回路SW1の端子a側に得られるオーディオ信号は、音響信号記録部12を介して記録媒体13、例えばメモ리카ード、磁気ディスクや光磁気ディスクなどの光ディスクや、ハードディスクに記録される。音響信号記録部12は、キー操作部20を通じて記録指示が使用者から与えられたときに、コントロール部10からの制御信号により、アクティブになり、記録を実行する。

入力端子 11 を通じた入力音響信号は、また、スイッチ回路 SW 2 の入力端 IN に供給される。このスイッチ回路 SW 2 の後段には、オーディオアンプ 15 及びスピーカ 16 が接続されており、後述する再生時以外は、スイッチ回路 SW 2 はコントロール部 10 からの切り換え制御信号により、入力端 IN 側に切り換えられる。したがって、再生時以外の時には、入力音響信号がオーディオアンプ 15 を通じてスピーカ 16 に供給されて、音響再生される。

キー操作部 20 を通じて再生指示が使用者から与えられると、コントロール部 10 からの制御信号により、音響信号再生部 14 がアクティブになり、記録媒体 13 から再生音響信号が読み出されて再生され、スイッチ回路 SW 2 の入力端 PB に供給される。この再生時には、スイッチ回路 SW 2 は、コントロール部 10 からの切り換え制御信号により、入力端 PB 側に切り換えられる。したがって、再生音響信号が、オーディオアンプ 15 を通じてスピーカ 16 に供給されて、音響再生される。

スイッチ回路 SW 1 の端子 a 側に得られる入力音響信号は、また、スイッチ回路 SW 3 の一方の入力端 d に供給される。このスイッチ回路 SW 3 の他方の入力端 c には、音響信号再生部 14 からの再生音響信号が供給される。

スイッチ回路 SW 3 は、検索対象の音響信号を切り換えるもので、キー操作部 20 を通じた使用者による検索対象の音響信号の選択指示に応じて切り換えられる。

このスイッチ回路 SW 3 から得られる検索対象の音響信号は、音響特徴パラメータ解析部 21 に供給されて、例えば、発音体の構造や共振系などの違いなどの音源の構造的な要素に依存する音響特徴

パラメータに解析する。この音響特徴パラメータとしては、ピッチ周波数、その高調波成分周波数、フォルマント周波数や、ケプストラム情報、線形予測係数とその残差信号、パーコール係数などが挙げられる。

例えば、図2は一般的な男性音声の周波数成分分布（図2（A））と時間波形（図2（B））の例である。また、図3は一般的な女性音声の周波数成分分布（図3（A））と時間波形（図3（B））の例である。また、図4はサキソホンの周波数成分分布（図4（A））と時間波形（図4（B））の例である。さらに、図5はベースギターの周波数成分分布（図5（A））と時間波形（図5（B））の例である。

これらの図2～図5は、単なる周波数成分分布及び時間波形であるが、それぞれ、ピッチ周波数、その高調波成分、フォルマント構造やその有無に、音響的な特徴が見られることが判る。

このような周波数成分分布の他にも、線形予測係数とその残差信号などの音響特徴パラメータでも、音響信号の特徴を表すことができる。

また、ピッチ周波数の時間的変動から得られる音声の抑揚パラメータも音響信号の特徴を表すパラメータとして有効であることが知られており、同一人物が話した言葉でも、抑揚によって別な状況と識別することができる。例えば、ピッチが通常より低く、さらに抑揚も下がる傾向のときには、元気のない場面や悲しい場面であると識別される。したがって、抑揚情報も音響特徴パラメータに用いることもできる。

この音響特徴パラメータ解析部21からの音響特徴パラメータは、

類似性算出部 22 に供給される。この類似性算出部 22 では、後述するように、検索子となる音響信号についての音響特徴パラメータ（以下、検索子音響特徴パラメータという）と、音響特徴パラメータ解析部 21 からの検索対象の音響信号の音響特徴パラメータとの類似性を算出する。

この類似性算出部 22 での類似性算出は、音声認識などのような音声時系列波形パターンの類似判定ではなく、発音体の構造や共振系について類似した構造を持った音源から発したものであるかの判定である。時系列波形のパターンの一致や類似度を算出するものではないので、例えばオーディオ信号の再生時に、再生速度を高速にしても類似性の算出が可能である。

すなわち、例えば、類似性の算出に当たって、音響特徴パラメータとして、前述したもののうちの、例えば、ピッチ周波数、フォルマント周波数などの 3 項目を取り上げて、パラメータ（ $x$ ， $y$ ， $z$ ）で表すことにする。

今、ある時間幅  $TW$  の区間での検索子音響特徴パラメータが、（ $x_0$ ， $y_0$ ， $z_0$ ）であり、検索対象の音響信号の音響特徴パラメータが、（ $x_i$ ， $y_i$ ， $z_i$ ）であったとしたとき、類似性算出部 22 では、次式に示す両者の距離  $S$  を類似度として算出する。前記時間幅  $TW$  は、類似性算出部 22 で、類似性を判定するのに十分な程度のパラメータ情報と時間を見込んで、予め設定される。

$$S = \sqrt{\{ (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 \}}$$

ここで、 $\sqrt{()}$  は、 $()$  内の平方根演算を示すものである。

この距離  $S$  が、小さければ小さいほど、類似度が大きいことにな

る。この実施の形態において、類似性算出部 22 では、予め定めた時間幅  $TW$  の区間ごとに、前記距離  $S$  を算出し、その算出した距離  $S$  が、予め定めた所定値  $S_{th}$  よりも小さいときに、前記時間幅  $TW$  の区間の音響信号は、検索子の音響信号と同じあるいは類似していると判定し、そうでないときには、非類似であると判定するようにしている。

なお、距離  $S$  の値は、類似の程度を表しているので、その類似の程度をそのまま出力として用いるようにすることも、勿論可能である。

この類似性算出部 22 の類似性判定出力は、表示制御部 23 を通じて表示部 24 に供給される。表示制御部 23 は、この実施の形態では、検索対象の音響信号の先頭からの時間に対応させて、前記類似性判定出力を、例えば横バーグラフのように、表示部 24 の画面に表示するようにする（後述する図 8 等参照）。

次に、検索子について説明する。この実施の形態では、検索子とする音響信号は、入力音響信号あるいは再生音響信号から抽出して、予め記憶部 32 に記憶させておく。

すなわち、スイッチ回路  $SW1$  の端子  $b$  に得られる入力音響信号は、スイッチ回路  $SW4$  の入力端  $f$  に供給され、また、音響信号再生部 14 からの再生音響信号が、スイッチ回路  $SW4$  の入力端  $e$  に供給される。このスイッチ回路  $SW4$  は、検索子の音響信号を入力音響信号から抽出するか、再生音響信号から抽出するかを、切り換えるもので、コントロール部 10 からの制御信号により、キー操作部 20 を通じた使用者の選択指示に応じて切り換えられる。

このスイッチ回路  $SW4$  からの音響信号は、検索子音響信号抽出

格納部 31 に供給される。この検索子音響信号抽出格納部 31 には、コントロール部 10 から、キー操作部 20 を通じた使用者の操作に応じた制御信号が供給される。

この実施の形態においては、使用者は、スピーカからの入力音響信号又は再生音響信号の音響再生出力音を聴取しながら、検索子としたい音響信号部分になったときに、キー操作部 20 の検索子指示キー（図示せず）を操作する。すると、そのキー操作時点を含む予め定めた時間幅  $TW$  の区間の音響信号を抽出し、それを検索子メモリ 32 に格納する。検索子メモリ 32 は、検索子単位の格納エリアを有し、各検索子は、それぞれ別々の格納エリアに格納される。

この例の場合、時間幅  $TW$  の区間は、キー操作時点よりもわずかに前の時点から開始するように、検索子音響信号抽出格納部 31 は、バッファメモリを備えている。そして、指示された時間幅  $TW$  の区間の音響信号を検索子音響信号として、検索子メモリ 32 に格納する。

この場合、検索子は、複数個が格納可能とされており、使用者が検索子の音響信号を指示するごとに、コントロール部 10 からの制御信号により、検索子メモリ 32 の検索子格納エリアが更新される。つまり、検索子格納エリアアドレスが更新される。そして、検索子音響信号抽出格納部 31 からの、使用者の指示により抽出された時間幅  $TW$  の区間の検索子の音響信号は、コントロール部 10 からの切り換え制御信号に応じて、検索子メモリ 32 の検索子格納エリアの更新に同期してスイッチ回路  $SW5$  が切り換えられることにより、順次に、検索子メモリ 32 の別々の検索子格納エリアに格納される。

このとき、図示しないが、どのような検索子の音響信号を検索子

メモリ 32 に格納したかを、キー操作部 20 を通じてメモ入力することができるようになっており、その検索子のメモ入力が、検索子格納エリア番号に対応して、例えばテーブル形式で表示部 24 に表示される。使用者は、この表示部 24 の検索子のテーブルを参照しながら、検索したい検索子が格納されている検索子格納エリア番号を指定することにより、所望の検索子により検索が実行できる。

すなわち、キー操作部 20 を通じて、使用者が検索モードを選択し、所定の検索子格納エリア番号を指定すると、コントロール部 10 からの制御信号により、スイッチ回路 SW6 が切り換えられるとともに、検索子音響信号読出部 33 により、検索子メモリ 32 の指定された検索子格納エリアから検索子の音響信号が読み出される。

読み出された検索子の音響信号は、検索子音響特徴パラメータ解析部 34 に供給されて解析され、これより、当該検索子の音響信号についての検索子音響特徴パラメータが得られる。そして、この検索子音響特徴パラメータが前述した類似性算出部 22 に供給され、前述したようにして、前記時間幅 TW の区間ごとに、検索対象の音響信号について、当該検索子の音響信号の検索が行われる。

以上のような構成の音響信号検索処理装置における動作について、さらに手順を追って説明する。

#### [検索子登録動作]

この実施の形態の場合、実際の検索処理に先立ち、検索子の登録が行われる。この例では、検索子は、前述したように音響信号の形で検索子メモリ 32 に格納されて登録される。この検索子登録のときのコントロール部 10 を主体にした動作のフローチャートは図 6 のようになる。この図 6 の処理ルーチンは、使用者がキー操作部 2

0 を通じて検索子登録モードを選択したときに開始となる。以下、この図 6 を参照して、検索子登録動作について説明する。

まず、コントロール部 10 は、キー操作部 20 を通じての使用者の選択指示を判定し、入力音響信号から検索子を抽出するのか、再生音響信号から検索子を抽出するのか判定する（ステップ S 101）。

この判定の結果、抽出元が、入力音響信号であると判別したときには、コントロール部 10 は、スイッチ回路 SW 1 は出力端 b 側に、スイッチ回路 SW 2 は入力端 IN 側に、スイッチ回路 SW 4 は入力端 f 側に、それぞれ切り換える（ステップ S 102）。

また、抽出元が再生音響信号であると判別したときには、コントロール部 10 は、スイッチ回路 SW 1 は出力端 a 側に、スイッチ回路 SW 2 は入力端 PB 側に、スイッチ回路 SW 4 は入力端 e 側に、それぞれ切り換える（ステップ S 103）。

ステップ S 102 又はステップ S 103 の次には、ステップ S 104 に進み、コントロール部 10 は、使用者によるキー操作部 20 を通じた検索子の音響信号部分の指示を待つ。このとき、検索子の抽出元のオーディオ信号は、オーディオアンプ 15 を通じてスピーカ 16 に供給されて、音響再生される。使用者は、そのスピーカ 16 から放音された音を聴取して、検索子としたい、例えば特定話者など特定の検索子音源として利用したい音響信号部分をキー操作部 20 を通じて指示入力する。

そして、ステップ S 104 で、検索子の音響部分の指示入力を検出したら、コントロール部 10 は、検索子メモリ 32 の検索子格納エリアを更新し（ステップ S 105）、また、検索子音響信号抽出

格納部 31 を制御して、前述したように、一定時間幅  $TW$  の区間の音響信号を抽出するようにする（ステップ S106）。そして、抽出した検索子の音響信号部分を、検索子メモリ 32 の、前記の更新した検索子格納エリアに書き込むように制御する（ステップ S107）。

そして、次に、キー操作部 20 を通じて検索子の登録モードの終了指示があったかどうか判別し（ステップ S108）、終了指示あれば、この検索子の登録モードの処理ルーチンを終了させ、終了指示がなければ、ステップ S104 に戻り、使用者による次の検索子の抽出指示入力を待ち、上述の動作を繰り返す。

以上のようにして、検索子メモリ 32 には、使用者が入力音響信号あるいは再生音響信号から抽出した検索子の音響信号を、登録して格納しておくことができる。この場合、検索子メモリ 32 のメモリ容量が十分に余裕がある場合には、検索子の音響信号は、消去せずに、蓄積しておくことができる。

この発明では、音響特徴パラメータを用いて類似性を算出するものであるので、蓄積した検索子の音響信号は、これを抽出したオーディオ信号から、当該検索子の音響信号部分と類似する部分を検索する場合にのみ用いられるのではなく、これを抽出したものではない他の検索対象の音響信号について、その検索子の音響信号部分と類似する部分を検索する場合にも用いることができる。

したがって、この第 1 図の実施の形態の場合、予め、使用者が検索したい音源についての音響信号を、前記時間幅  $TW$  ずつ検索子メモリ 32 に格納して蓄積しておくことにより、何時でも必要なときに、その検索子を用いた検索を行うことができる。

### [検索処理動作]

以上のようにして検索子メモリ 32 に格納された検索子の音響信号を用いて、次のようにして検索処理が実行される。この検索処理の処理ルーチンを図 7 に示す。この図 7 の処理ルーチンは、使用者がキー操作部 20 を通じて検索処理モードを選択したときに開始となる。以下、この図 7 を参照して、検索処理動作について説明する。

まず、コントロール部 10 は、キー操作部 20 を通じての使用者の選択指示を判定し、検索対象は、入力音響信号か、再生音響信号かを判定する（ステップ S 201）。

この判定の結果、検索対象が入力音響信号であると判別したときには、コントロール部 10 は、スイッチ回路 S W 1 は出力端 a 側に、スイッチ回路 S W 2 は入力端 I N 側に、スイッチ回路 S W 3 は入力端 d 側に、それぞれ切り換える（ステップ S 202）。

また、検索対象が再生音響信号であると判別したときには、コントロール部 10 は、スイッチ回路 S W 2 は入力 P B 側に、スイッチ回路 S W 3 は入力端 c 側に、それぞれ切り換える（ステップ S 203）。

ステップ S 202 又はステップ S 203 の次には、ステップ S 204 に進み、コントロール部 10 は、使用者により検索子として何が選択されたかを認識し（ステップ S 204）、その選択された検索子の音響信号を、検索子メモリ 32 から読み出すように制御する（ステップ S 205）。

すると、メモリ 32 から読み出された検索子の音響信号は、検索子音響信号読出部 33 を介して検索子音響特徴パラメータ解析部 34 に供給されて、検索子音響特徴パラメータとされて、類似性算出

部 2 2 に供給される。類似性算出部 2 2 では、検索対象のオーディオ信号の例えば時間幅  $TW$  の区間ごとの音響特徴パラメータと、検索子音響特徴パラメータとの前述したような類似性演算がなされて、両者の類似性の指標である類似度が算出される（ステップ S 2 0 6）。この例では、前述したように、類似度は、両者の距離  $S$  として求められる。

そして、この類似度である前記距離  $S$  が、予め定めた閾値  $S_{th}$  よりも小さいか否かを判別し（ステップ S 2 0 7）、閾値  $S_{th}$  よりも小さいときには、検索子に類似する音響信号部分として判定し（ステップ S 2 0 8）、閾値  $S_{th}$  よりも大きいときには、非類似であると判定する（ステップ S 2 0 9）。そして、その判定結果を、表示制御部 2 3 を介して表示部 2 4 の画面に表示する（ステップ S 2 1 0）。

そして、検索対象のオーディオ信号が終了することにより、あるいは、使用者の終了指示により、検索処理が終了したか否かを判定し（ステップ S 2 1 1）、終了であれば、この処理ルーチンを終了し、そうでなければ、ステップ S 2 0 6 に戻って、検索子による検索を継続する。

以上のようにして、検索された結果の表示例を図 8 に示す。この図 8 の例は、検索対象が会議録音の再生音響信号であって、検索子が特定の話者 A である場合である。この表示例によれば、会議録音全体の中のどの時間的位置で話者 A が発言しているかが示されることになる。

なお、以上の例では、検索子は、一つとして説明したが、複数個の検索子を指定し、その複数個の検索子についての類似性を算出し

て、表示するようにすることもできる。その場合、例えば10数ミリ秒ごとに、検索子を時分割的に切り換えて、類似性算出部22で算出演算を行うようにすればよい。

また、類似性算出部を複数個設け、それらの類似性算出部で並列的に複数個の検索子のそれぞれについての類似性算出を行うようにしてもよい。

そのように複数個の検索子を設定して検索を行った検索結果の表示例を図9に示す。この表示例は、会議録音全体の中で、3名の話者A、B、Cがそれぞれどのような時間区間で発言を行ったかを示している。

こうして、この実施の形態によれば、例えば、複数の会議参加者が参加している会議において、各会議参加者ごとの音響信号を検索子として予め登録しておけば、会議の時間経過に伴った各会議参加者ごとの発言区間を分類して表示することが可能になる。

なお、以上の説明では、類似性の判定結果は、所定の閾値を超えたか否かによる2値的なものとしたが、前述の例の場合には、類似性の算出結果としての距離Sの値をそのまま表示に反映するようにしてもよい。すなわち、距離Sが短ければ短いほど類似度が高いとして、その類似度の高さに応じて、図8(B)に示すように、時間区間表示に濃淡を付けて、表示するようにしてもよい。例えば、図8(B)の例では、濃淡が濃い区間ほど、類似性が高いことを示すようにしている。また、濃淡の代わりに、類似度の高さの違いを表示色の違いにより、表示するようにしてもよい。

また、検索子が複数の場合には、検索子ごとに類似性判定出力の表示色を異ならせ、また、類似度の高さの違いを表示色の濃度の違

いにより、表示するようにしてもよい。

なお、以上の表示例は、類似性判定出力を検索対象の音響信号の時間経過に対応して表示したものである。一方、検索対象の音響信号が再生音響信号の場合には、各時間区間の音響信号の記録媒体13上における記録位置は容易に知ることができる。したがって、一度、再生音響信号についての検索結果としての類似性判定表示出力を得たら、それを表示部23に表示しながら、図10に示すように、その類似性判定表示出力に現在再生位置を、例えば再生位置カーソルによって対応表示するようにすることもできる。

また、逆に、図10の再生位置カーソルを、使用者が指定できるようにすることにより、その指定した再生位置カーソルの位置から検索対象の音響信号の再生を行うようにすることもできる。その場合には、検索により、分類した検索子ごとの音響信号を、検索対象の音響信号から抽出して再生することが可能になる。

なお、検索対象の音響信号と検索子との類似性を算出するのは、この発明においては、音響特徴パラメータを用いているので、波形パターンが一致するかどうかのパターンマッチング処理を用いる音声認識などとは異なり、音響信号の時系列波形をそのまま用いる必要がない。このため、再生音響信号を検索対象とする場合、再生速度はノーマル速度（1倍速）ではなく、2倍速や3倍速というような高速再生を行っても、同様の類似性判定出力を得ることができる。したがって、検索速度を高速にすることが可能である。

したがって、検索子を用いた検索を高速再生速度にて行い、検索子と類似している音響信号部分を検出したときに、その部分をノーマル再生するようにすることにより、検索子に類似する部分のみを

抽出再生することができる。

また、一旦、検索を行えば、記録媒体上のアドレスと、検索結果とを対応させることができるので、その検索結果を用いて、特定の検索子と類似する音響信号部分のみを記録媒体から抽出して、再生するようにすることも可能である。

以上の効果の説明は、検索対象が再生音響信号の場合であるが、この実施の形態では、入力音響信号を検索対象として、予め登録してある検索子を用いて検索して、入力音響信号について、検索子ごとの分類などを行うことができるという効果もある。例えば、複数個の楽器の音響信号を検索子として登録しておき、入力音響信号の各パートが、どのような楽器を用いて演奏されたものであるかなどを分類表示することができる。

なお、以上の実施の形態では、検索子は、入力音響信号や再生音響信号から抽出して、検索子メモリ 32 に格納するようにしたが、検索子ライブラリとして複数個の検索子の音響信号を検索子メモリ 32 に予め記憶しておいても、もちろんよい。

また、類似性の算出は、音響特徴パラメータ間の幾何学的距離  $S$  として求めるようにしたが、幾つかのパラメータから新しいパラメータを算出し、それを用いて音響特徴パラメータ間の距離を求めるようにすることもできる。

例えば、上述の例の  $x$ ,  $y$ ,  $z$  の 3 項目の音響特徴パラメータを用いる場合において、新しいパラメータを  $k$  としたとき、

$$k = a y + b z$$

として、パラメータ  $x$  と、この新たなパラメータ  $k$  との 2 つのパラメータ ( $x$ ,  $k$ ) を用いて、検索対象と検索子の音響特徴パラメー

タ間の距離を算出するようにしてもよい。また、それぞれの音響特徴パラメータに重み付けを行って、重要なパラメータに着目した類似性を算出するようにすることもできる。

また、検索子メモリ32は、内蔵のメモリではなく、メモリカードなどの取り出し可能のものであってもよい。その場合には、複数個のメモリカードに、複数種の検索子を記憶して用意しておき、メモリカードを検索対象のオーディオ信号に合わせて交換することにより、希望する検索対象について、希望の検索子を用いた検索を容易に行うことができる。

また、検索子とする音響信号の音源を用意し、その音源からの音響信号を予め入力端子11から入力し、検索子メモリ32に、検索子の音響信号として記憶格納するようにすることもできる。さらに、検索子メモリ32に記憶することなく、検索子の音源からの音響信号を検索対象の音響信号とは別に入力し、バッファメモリにその検索子の音響信号を一時的に蓄えて、検索に用いるようにしてもよい。

なお、検索子は、単一音源からの音響信号に限定されるものではなく、入力音響信号や再生音響信号から抽出された所定区間の音響信号であってもよいことはいうまでもない。

なお、表示部には、画像を伴う音響信号の場合には、画像も一緒に表示することもできる。画像を併せて表示することで、検索を行った結果の分類の確認をすることができる。例えば、録画内容を音響信号を使用して検索、分類することも可能である。

#### [第2の実施の形態]

図11に、この第2の実施の形態の音響信号検索処理装置のプロ

ック図を示す。第 1 の実施の形態は、登録検索子として、音響信号を格納するようにしたが、この第 2 の実施の形態では、登録検索子として、その音響特徴パラメータを検索子メモリ 32 に格納するようにする。その他の構成は、図 1 の装置と全く同様であり、図 1 と同一部分については、同一符号を付してある。

すなわち、この第 2 の実施の形態においては、スイッチ回路 SW 4 の出力側には、検索子音響信号抽出部 35 が設けられる。これは、図 1 の検索子音響信号抽出格納部 31 のうちの、使用者の指示した検索子の音響信号部分を抽出する手段の部分に相当する。

この検索子音響信号抽出部 35 で抽出された検索子の音響信号は、検索子音響特徴パラメータ解析部 36 で解析されて、音響特徴パラメータとされ、検索子音響特徴パラメータ格納部 37 及びスイッチ回路 SW 5 を通じて検索子メモリ 32 に格納される。検索子音響特徴パラメータ格納部 37 は、格納する対象情報が、検索子音響特徴パラメータに変わるだけで、図 1 の検索子音響信号抽出格納部 31 のうちの、検索子として抽出した音響信号部分を検索子メモリ 32 に格納する手段に相当し、前述と同様にして、各検索子（この実施の形態では、音響特徴パラメータ）を検索子メモリ 32 の別々の検索子格納エリアに書き込む。

検索子メモリ 32 に格納された検索子の音響特徴パラメータは、キー操作部 20 を通じた使用者の指示に基づいて生成されるコントロール部 10 からの制御信号により、スイッチ回路 SW 6 及び検索子音響特徴パラメータ読出部 38 によって読み出される。この第 2 の実施の形態の場合には、読み出されたものは音響特徴パラメータであるので、そのまま、類似性算出部 22 に供給されて、検索のた

めの類似性算出が実行される。

この第2の実施の形態は、検索子メモリ32に検索子音響特徴パラメータが格納される点が異なるのみで、検索子の登録動作や、検索子を用いた検索動作は、前述した第1の実施の形態と同様に行われるものである。したがって、この第2の実施の形態によれば、上述した第1の実施の形態の場合と同様の作用効果が得られる。また、上述した第1の実施の形態と同様の変形例が可能である。

さらに、この第2の実施の形態では、検索子としては、その音響信号そのものではなく、検索子音響特徴パラメータを検索子メモリ32に記憶するので、一つの検索子当たりの記憶容量が少なくて済み、格納して蓄積することができる検索子の数を多くすることが可能である。

#### [第3の実施の形態]

この第3の実施の形態は、音響特徴パラメータとして、音響信号への再合成が可能なものを用いる場合である。音響特徴パラメータから音響信号への再合成が可能であることから、音響信号の記録も、音響特徴パラメータとして記録するようにする。そして、この第3の実施の形態では、第2の実施の形態と同様に、検索子は、その音響特徴パラメータを検索子メモリ32に記憶格納しておくようにする。

図12は、この第3の実施の形態の音響信号検索処理装置のブロック図である。この図12においても、前述の実施の形態の場合と同一部分には同一符号を付すことにする。この第3の実施の形態では、音響特徴パラメータとして、例えばボコーダのパラメータ等を用いる。



図 1 2 に示すように、この第 3 の実施の形態の場合には、入力端子 1 1 を通じて入力音響信号は、音響特徴パラメータ解析部 4 1 に供給されて、例えばボコーダ方式の音響特徴パラメータに解析される。そして、この音響特徴パラメータ解析部 4 1 からの音響特徴パラメータが、スイッチ回路 S W 1 を通じて音響特徴パラメータ記録部 4 2 に供給され、コントロール部 1 0 の制御の下、この音響特徴パラメータ記録部 4 2 により、入力音響信号の音響特徴パラメータが記録媒体 1 3 に記録される。

そして、記録媒体 1 3 に記録された入力音響信号の音響特徴パラメータは、コントロール部 1 0 の制御の下、音響特徴パラメータ再生部 4 3 により再生され、音響信号合成部 4 4 に供給される。この音響信号合成部 4 4 は、音響特徴パラメータから、元の音響信号を合成する。

この音響信号合成部 4 4 から得られた再生音響信号は、スイッチ回路 S W 2 の入力端 P B 側を通じ、オーディオアンプ 1 5 を通じてスピーカ 1 6 に供給され、音響再生される。なお、入力端子 1 1 からの入力音響信号がスイッチ回路 S W 2 の入力端 I N 側及びオーディオアンプ 1 5 を通じてスピーカ 1 6 に供給されて、音響再生されるのは、第 1 及び第 2 の実施の形態の場合と同様である。

そして、この第 3 の実施の形態では、音響特徴パラメータ解析部 2 1 は設けられず、スイッチ回路 S W 3 は、スイッチ回路 S W 1 の出力端 a 側を通じて入力音響信号の音響特徴パラメータと、音響特徴パラメータ再生部 4 3 からの再生音響特徴パラメータとのいずれかを、キー操作部 2 0 を通じた使用者の指示に応じたコントロール部 1 0 からの切り換え制御信号より選択して、類似性算出部 2 2 に

供給する。

一方、検索子は、この第3の実施の形態では、第2の実施の形態と同様に、検索子音響特徴パラメータを検索子メモリ32に格納する。このため、この第3の実施の形態の場合には、音響特徴パラメータ解析部41からの音響特徴パラメータが、スイッチ回路SW1の出力端b側を通じてスイッチ回路SW4の一方の入力端fに供給され、また、音響特徴パラメータ再生部43からの再生音響パラメータがスイッチ回路SW4の他方の入力端eに供給される。

そして、スイッチ回路SW4がコントロール部10からの切り換え制御信号により、検索子の音響特徴パラメータを入力音響信号から抽出するか、再生音響信号から抽出するかに応じて、入力端e又は入力端fのいずれかに切り換えられる。

そして、このスイッチ回路SW4からの音響特徴パラメータが検索子音響特徴パラメータ抽出格納部45に供給される。前述の第1の実施の形態の場合と同様にして、使用者は、スピーカからの入力音響信号又は再生音響信号の音響再生出力を聴取しながら、検索子としたい音響信号部分になったときに、キー操作部20の検索子指示キー（図示せず）を操作する。すると、検索子音響特徴パラメータ抽出格納部45は、そのキー操作時点を含む予め定めた時間幅TWの区間の音響信号の音響特徴パラメータを抽出し、検索子メモリ32の検索子格納部に格納する。

こうして、検索子メモリ32には、入力音響信号あるいは再生音響信号のうちから抽出された検索子の音響信号の音響特徴パラメータが格納される。この検索子メモリ32に格納された検索子音響特徴パラメータは、前述の第2の実施の形態と同様にして、使用者の

指示に応じて、検索子音響特徴パラメータ読出部38により読み出されて、類似性算出部22に供給される。

検索子の登録動作と、検索動作は、前述の2つの実施の形態とほぼ同様であるので、その詳細は省略する。このように、この第3の実施の形態も、検索子メモリ32に検索子音響特徴パラメータが格納される点が異なるのみで、検索子の登録動作や検索動作は、前述した第1の実施の形態や第2の実施の形態と同様に行われるものである。したがって、この第3の実施の形態の場合も、上述した第1の実施の形態の場合と同様の作用効果が得られ、また、上述した第1の実施の形態と同様の変形例が可能である。

そして、この第3の実施の形態の場合には、特に、音響特徴パラメータを音響信号の代わりに記録媒体に記録するので、記録媒体の記録容量を有効に活用することができる。そして、第1の実施の形態や第2の実施の形態に比べて、音響特徴パラメータ解析部を、検索対象と検索子とで共通の一つとすることができて、構成を簡略化することができる。

再合成可能な音響特徴パラメータとしての具体的な例としては、MPEG (Moving Picture Experts Group) 規格の音響圧縮形式 (MPEGオーディオのレイヤI、レイヤII、レイヤIII) データやATRAC (Adaptive Transform Acoustic Coding) 形式のデータなどを用いる例も挙げられる。

これらMPEGオーディオ規格やATRACでは、人間の聴覚心理特性を効率良く利用するために、オーディオ帯域を複数個に細分化してサブバンド符号化を行っている。例えばMPEGオーディオ

のレイヤ I では、オーディオ信号の全帯域を等間隔の周波数幅で 32 個のサブバンドに分割し、それぞれの信号を元のサンプリング周波数の  $1/32$  でサブサンプリングして符号化する。

そこで、MPEGオーディオ規格の音響圧縮形式のデータやATRAC形式のデータ（音響圧縮符号化データ）を検索対象とする場合には、そのサブバンドデータから、帯域レベル分布の算出、主要帯域周波数の算出、調波構造の検出、主要帯域周波数の時間推移の検出などを行い、その検出結果を検索子音響特徴パラメータとする。

MPEGオーディオ規格の音響圧縮形式のデータやATRAC形式のデータを用いる方法を、図12の例に適用するときは、入力端子11からの入力音響信号がアナログ信号やリニアPCM信号の場合には、音響特徴パラメータ解析部41では、前記MPEG規格の音響圧縮形式データやATRAC形式のデータを生成する。そして、記録部42は、その音響圧縮データを記録媒体13に記録するようにする。

そして、検索子音響特徴パラメータ抽出格納部45では、音響特徴パラメータ解析部41からの、あるいは再生部43からの音響圧縮データからサブバンドのデータを生成し、そのサブバンドのデータについて、前述のように、帯域レベル分布の算出、主要帯域周波数の算出、調波構造の検出、主要帯域周波数の時間推移の検出などを行い、その検出結果を検索子音響特徴パラメータとして、検索子メモリ32に書き込むようにする。

そして、この場合、類似性算出部22の前段には、図1や図11の例と同様に、音響特徴パラメータ解析部を設け、スイッチSW1からの、あるいは再生部43からの音響圧縮データからサブバンド



のデータを生成し、そのサブバンドのデータについて、前述のように、帯域レベル分布の算出、主要帯域周波数の算出、調波構造の検出、主要帯域周波数の時間推移の検出などを行うようにする。この場合の音響特徴パラメータ解析部は、構成が簡単になる。そして、類似性算出部 22 は、その解析により得た音響特徴パラメータと、検索子音響特徴パラメータ読出部 38 からの検索子音響特徴パラメータとを比較して、類比判定を行う。

以上の入力端子 11 からの入力音響信号がアナログ信号やリニア PCM 信号の場合であるが、IEEE (The Institute of Electrical and Electronics Engineers, Inc.) 1394 規格のインターフェースからのシリアルデジタルデータで入力された場合や、デジタル放送受信装置においてデジタル音響圧縮信号として得られる信号が入力される場合には、MP EG 規格の音響圧縮形式データや ATRAC 形式のデータがそのまま得られるので、図 12 の音響特徴パラメータ解析部 41 は不要となる。

以上説明した音響圧縮信号を検索対象とする検索方法は、デジタルオーディオ圧縮形式データを復号化して、リニア PCM オーディオ信号に戻してから音響特徴分析するよりも分析のための時間遅れが小さくなるとともに、演算規模が小さくなる、という利点がある。

#### [第 4 の実施の形態]

第 3 の実施の形態では、音響特徴パラメータを音響信号の代わりに記録媒体に記録するようにしたが、第 4 の実施の形態では、音響信号と、音響特徴パラメータとを、互いに付随的に記録媒体に記録するようにする。この第 4 の実施の形態の場合は、音響信号が音響

特徴パラメータとともに記録されるので、音響特徴パラメータは、第 3 の実施の形態のように音響信号に再合成できるものに限定されない。

そして、この第 4 の実施の形態においては、磁気ディスク、光磁気ディスク、光ディスク、メモリカードなどの記録媒体へは、音響信号に付随して音響特徴パラメータを記録するようにする。すなわち、音響信号と音響特徴パラメータとは、音響信号の記録単位区間（バック、パケット、セクタなどの区間）分ごとに記録媒体に記録するが、音響信号の記録単位区間分のデータと、対応する単位区間分の音響特徴パラメータは、隣接した記録エリア等、互いに対応がとれる形態で、つまり、リンクされた構造で記録される。

図 13 に、この第 4 の実施の形態の音響信号検索処理装置のブロック図を示す。この図 13 において、前述の実施の形態の図 1、図 11 及び図 12 に示した各部と同一部分には同一符号を付して、その詳細な説明は省略する。

すなわち、この第 4 の実施の形態では、オーディオ入力端子 11 を通じた入力音響信号は、スイッチ回路 SW7 を通じて音響信号記録部 12 に供給される。スイッチ回路 SW7 は、コントロール部 10 からの切り換え制御信号により、記録時にオンとされる。そして、この第 4 の実施の形態では、音響信号記録部 12 を通じた入力音響信号は、記録データ制御部 51 に供給される。

また、入力端子 11 を通じた入力音響信号は、音響特徴パラメータ解析部 41 に供給されて、音響特徴パラメータに解析される。この音響特徴パラメータ解析部 41 からの音響特徴パラメータは、検索対象の入力音響信号の音響特徴パラメータとしてスイッチ回路 S

W 3 の入力端 d に供給されるとともに、入力音響信号から検索子の音響特徴パラメータを抽出するために、スイッチ回路 S W 4 の入力端 f に供給される。さらに、音響特徴パラメータ解析部 4 1 からの音響特徴パラメータは、記録データ制御部 5 1 に供給される。

記録データ制御部 5 1 は、コントロール部 1 0 の制御を受けて、記録時には、予め定められている記録フォーマットで、音響信号と音響特徴パラメータとを対応付けて記録する。この例では、記録データ制御部 5 1 は、記録時には、前述したように、音響信号を記録単位区間分ごとに、記録媒体 1 3 に記録するとともに、対応する記録単位区間分の音響特徴パラメータを、記録媒体 1 3 の、当該音響信号記録位置に隣接する位置に記録する。

そして、上記のような記録フォーマットで記録媒体 1 3 に記録された入力音響信号及びその音響特徴パラメータは、読み出しデータ制御部 5 2 により、コントロール部 1 0 の制御を受けて再生される。そして、読み出しデータ制御部 5 2 で記録媒体 1 3 から再生された音響信号は、音響信号再生部 1 4 に供給され、スイッチ回路 S W 2 の入力端 P B 側を通じ、オーディオアンプ 1 5 を通じてスピーカ 1 6 に供給されて、音響再生される。

また、読み出しデータ制御部 5 2 で記録媒体 1 3 から再生された対応する音響特徴パラメータは、検索対象の再生音響信号の音響特徴パラメータとして、スイッチ回路 S W 3 の入力端 c に供給されるとともに、再生音響信号から検索子の音響特徴パラメータを抽出するために、スイッチ回路 S W 4 の入力端 e に供給される。

さらに、読み出しデータ制御部 5 2 で記録媒体 1 3 から再生された音響信号及び音響特徴パラメータは、他の装置で音響信号につい

ての検索を行うために、外部出力端子（図示せず）を通じて、外部へ転送可能とされている。したがって、検索を行う他の装置では、音響特徴パラメータを作成するために、音響特徴パラメータ解析部を備える必要はない。

この第4の実施の形態は、入力音響信号と、その音響特徴パラメータとが、リンクさせた状態で、記録データ制御部51により、記録媒体に記録され、また、互いにリンクされて、音響信号と、その音響特徴パラメータとが、読み出しデータ制御部52により読み出されて再生される点が、前述の実施の形態と異なるが、音響特徴パラメータを用いて検索を行い、また、入力音響信号や再生音響信号について検索子音響特徴パラメータの抽出を行えるようにした点は、第3の実施の形態と同様である。

したがって、この第4の実施の形態の場合も、上述した第1の実施の形態の場合と同様の作用効果が得られ、また、上述した第1の実施の形態と同様の変形例が可能である。

そして、この第4の実施の形態の場合には、第3の実施の形態と同様に、第1の実施の形態や第2の実施の形態に比べて、音響特徴パラメータ解析部を、検索対象と検索子とで共通の一つとすることができて、構成を簡略化することができる。また、他の検索装置で検索を行う場合にも、当該他の検索装置に、音響特徴パラメータを供給することができるので、他の検索装置には、音響特徴パラメータ解析部は設ける必要がないという利点がある。

なお、記録する音響信号と、対応する音響特徴パラメータとは、リンクして記録媒体に記録されていればよいので、上述の例のように必ずしも隣接して記録される必要はなく、記録媒体に、予め音響

信号の記録エリアと音響特徴パラメータの記録エリアを別々に決めておき、それぞれの記録エリアの対応関係にある（リンクされている）アドレスに、対応する音響信号と音響特徴パラメータのそれぞれを記録するようにしてもよい。また、音響信号と音響特徴パラメータのリンク情報を別に記録して、そのリンク情報を用いて、両者に対応付けて、再生するようにしても勿論よい。

#### 〔第 5 の実施の形態〕

上述したように、この発明によれば、検索対象の音響信号について、音響特徴パラメータによって、検索子と類似性判定をすることにより、検索が可能であるので、検索対象の音響信号を記録しながら、指定した検索子についての検索を並行して行うことができる。したがって、記録終了時には、当該検索対象の音響信号についてのその検索子についての検索結果が得られる。そこで、記録媒体には、音響信号の記録領域と別個のエリアに、この検索結果を合わせて記録することも可能である。

また、再生音響信号についての検索結果も、音響信号の記録領域と別個のエリアを記録媒体に用意することにより、後から記録媒体に記録することも可能である。

第 5 の実施の形態においては、以上の点を考慮して、検索対象の音響信号とともに、その検索結果を記録するようにする。

すなわち、この第 5 の実施の形態では、予め登録した検索子の一つあるいは複数個を用いて、記録する入力音響信号についての検索を、記録処理に並行して行う。

例えば、記録媒体 13 としては、磁気ディスク、光磁気ディスク、光ディスクなどのディスク媒体やメモリカードを用いる。その場合、

記録時には、記録媒体上のアドレスを監視しながら、記録が行われる。この第5の実施の形態では、検索子による検索の結果、類似すると判定された音響信号部分が、記録媒体上のどのアドレス（記録位置）に記録されているかを、その検索子の識別情報に対応して記憶しておく。検索子の識別情報としては、例えば、この例では、スイッチ回路SW6の切り換え制御情報（検索子格納エリア、つまり検索子ナンバーに対応）を用いる。

図14は、この第5の実施の形態の場合において、記録終了後の、検索結果をテーブルとして表したものである。このテーブルは、記録媒体に記録された信号についての検索用インデックステーブルとなる。

この図14において、検索子名は、例えば、検索子の登録時に予め付与した名称であり、例えば、会議録音などの場合には、会議参加者名となる。区間データは、検索子と類似すると判定された区間を示すデータで、その区間の開始アドレスSTと、終了アドレスEDとで表される。また、ADR<sub>i</sub>（iは整数）は、記録媒体上のアドレス値をそれぞれ示すものである。

この第5の実施の形態においては、この検索用インデックステーブルの情報を、記録媒体上の予め定めた記録エリアに、記録音響信号とは別個に記録するようにする。

この第5の実施の形態の場合、この検索用インデックステーブルを用いることにより、例えば特定の検索子と類似する音響信号部分を抽出して再生（検索再生と称する）することが容易にできる。

図15は、その検索再生の際の処理ルーチンを示すものである。この処理ルーチンは、キー操作部を通じて、検索再生指示を行った

記入入力音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記検索対象の音響信号中から前記検索子の音響信号と類似する部分を検索し、その検索結果を、前記検索子及び前記入入力音響信号の前記記録媒体上の記録位置に関連させて記録するようにすることを特徴とする音響情報蓄積方法。

13. 前記検索結果は、前記音響信号とともに、前記記録媒体に記録することを特徴とする請求の範囲第12項に記載の音響情報蓄積方法。

14. 検索子とする1～複数の音響信号についての前記検索子音響特徴パラメータのそれぞれは、前記入入力音響信号から、使用者の指示に応じて選択抽出した音響信号部分についての解析結果として求められたものであることを特徴とする請求の範囲第12項又は請求項第13項のいずれか1項に記載の音響情報蓄積方法。

15. 前記使用者の指示に応じて選択抽出した音響信号部分についての解析結果として求められた前記検索子とする1～複数の音響信号についての前記検索子音響特徴パラメータのそれぞれは、バッファメモリの異なるメモリエリアに記憶され、前記バッファメモリから、前記検索子音響特徴パラメータのそれぞれが読み出されて、それらの検索子音響特徴パラメータを用いた前記入入力音響信号についての前記類似性の算出処理による前記類似する部分の検索処理が行われることを特徴とする請求の範囲第14項に記載の音響情報蓄積方法。

16. 前記入入力音響信号から、使用者の指示に応じて選択抽出した1～複数個の音響信号部分が、検索子の音響信号として、それぞれバッファメモリの異なるメモリエリアに記憶され、前記バッファ

する請求の範囲第1項に記載の音響検索処理方法。

6. 前記検索子音響特徴パラメータは、前記検索対象の音響信号中から抽出した前記検索子の音響信号を解析して得た音響特徴パラメータであることを特徴とする請求の範囲第5項に記載の音響検索処理方法。

7. 前記検索子が複数個である場合には、予め設定された単位時間ごとに、検索子を切り換えて、前記類似性の算出を行うことを特徴とする請求の範囲第1項乃至第5項のいずれか1項に記載の音響検索処理方法。

8. 前記検索子毎の検索結果を、前記検索対象の音響信号について、時間方向に表示することを特徴とする請求の範囲第1項乃至第6項のいずれか1項に記載の音響検索処理方法。

9. 前記検索子毎の検索結果を、前記検索対象の音響信号について、時間方向に表示することを特徴とする請求の範囲第7項に記載の音響検索処理方法。

10. 前記検索対象の音響信号のうち、前記検索の結果、前記検索子と類似する部分のみを抽出して再生することを特徴とする請求の範囲第1項乃至第6項のいずれか1項に記載の音響検索処理方法。

11. 前記検索対象の音響信号のうち、前記検索の結果、前記検索子と類似する部分のみを抽出して再生することを特徴とする請求の範囲第7項に記載の音響検索処理方法。

12. 入力音響信号を記録媒体に記録するとともに、前記入力音響信号を解析して、音源の構造的な要素に依存する音響特徴パラメータを得、検索子とする1～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前

### 請求の範囲

1. 検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記検索対象の音響信号中から前記検索子の音響信号と類似する部分を検索することを特徴とする音響検索処理方法。
2. 前記検索子とする1～複数の音響信号を用意し、検索時に、前記用意されている検索子の音響信号のうちの、検索子とすべき音源の音響信号を解析して得た音響特徴パラメータを、前記類似性の算出に用いる前記検索子音響特徴パラメータとすることを特徴とする請求の範囲第1項に記載の音響検索処理方法。
3. 前記検索子の音響信号は、前記検索対象の音響信号中から抽出して記憶しておくことを特徴とする請求の範囲第2項記載の音響検索処理方法。
4. 検索時に、前記検索対象の音響信号について使用者が指示操作をした部分を解析して得たものを前記検索子音響特徴パラメータとし、この検索子音響特徴パラメータを用いて、前記検索対象の音響信号についての検索を行うことを特徴とする請求の範囲第1項に記載の音響検索処理方法。
5. 予め、前記1～複数の検索子音響特徴パラメータを記憶しておき、検索時に、前記記憶されている検索子音響特徴パラメータの中から読み出したものを、前記類似性の算出に用いることを特徴と

響特徴パラメータと映像特徴パラメータとを用いてシーンの検索を行うようにすれば、検索の精度を上げることができ、また、音響的にも映像的にも区切りのよいシーン検索ができるようになる。

に、後の時点で読み出して再生するようにすることもできる。

この第2の実施の形態によれば、映像特徴量を使用してシーン区間の検出することと、音響特徴量を使用してシーン区間を検出することとを併用することにより、検索結果のシーン区間を意味あるものとすることができる。例えば、映像的にはシーンが代わっていても、ナレーションなど音響的には続いているシーンを、映像の変化点で区切られることなく、検出することができる。

また、前述の縮小静止画表示用の、縮小静止画として、検索子映像特徴パラメータと検索子音響特徴パラメータとにより検索された類似部分の重なり部分を抽出するようにすることにより、映像的だけでなく音響的にも意味のある縮小静止画を抽出して表示することができる。

以上説明したように、この発明によれば、検索対象の音響信号と、検索子としての音響信号とを、それぞれ音響特徴パラメータに解析し、その音響特徴パラメータについて比較して、類比判定することにより検索を行うようにしたことにより、検索対象の音響信号の記録時にマーカなどを打ち込んでおかなくても、例えば、特定の話者の発言部分や、特定の楽器の演奏部分を、簡単に検索することができる。

また、予め検索子を登録しておくことにより、その登録したものから選択した検索子についての検索子を容易に行うことができる。

また、マーカを施す場合のように、一旦記録しなければ検索ができないのではなく、予め用意した検索子についての検索を、検索対象の音響信号について行うことが可能である。

また、音響映像信号からの任意のシーンの検索を行う場合に、音

結果と、前記映像信号についての検索結果とに基づいて、入力音響映像信号中から所定の部分を検索するようにする。

この場合に、検索子映像特徴パラメータと、検索子音響特徴パラメータとは、それぞれ別々の検索子として登録して別々に用いるようにしても良いし、互いに関連付けて登録して用いるようにしても良い。例えば、ある人物の映像の検索子映像特徴パラメータと、その人物の声紋などの検索子音響特徴パラメータとは互いに関連付けて登録しておくが良い。

この第2の実施の形態においては、検索に当たっては、検索子映像特徴パラメータと、検索子音響特徴パラメータとの両方をそれぞれ個別に指定して検索を行うことができる。また、前記の互いに関連付けられている検索子映像特徴パラメータ及び検索子音響特徴パラメータを用いるように検索子の指定をすることもできる。

検索子音響特徴パラメータと検索子映像特徴パラメータの両者を用いた検索の方法としては、それぞれのパラメータのいずれかにより検索された類似部分の全てを検索結果として出力する方法、両パラメータで検索された類似部分のうち、互いに一部あるいは全部が重なり合う部分を検索結果として出力する方法などを指定することができる。

このときの検索結果は、検索された部分を、アドレス情報などで表示しても良いし、また、映像部分については、検索された部分の先頭のシーンを前記の縮小静止画にして表示するようにしてもよい。また、前述の音響検索処理の各実施の形態と同様に、検索結果の信号部分を順次に再生するようにすることもできる。さらに、検索結果と、検索子とを関連させてメモリに記憶しておき、検索子を頼り

提供された検索子の中から所望のものを選択して検索に利用したり、また、その提供されたものの中から必要な検索子音響信号や検索子音響特徴パラメータを検索子メモリ 32 に登録して保存するようにすることもできる。

〔音響映像検索装置、音響映像蓄積装置の第 2 の実施の形態〕

上述の第 1 の実施の形態では、音響信号についてのみ検索子を用いて検索を行うようにしたが、この第 2 の実施の形態では、映像信号についても、同様にして検索を行うようにする。

すなわち、この第 2 の実施の形態では、映像信号の検索のために、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータを登録しておくようにする。この検索子映像特徴パラメータとして抽出する映像の特徴量としては、例えば、輝度とその変化、色合いとその変化、人物、物体などの映っているものとその変化などを用いることができる。

この検索子映像特徴パラメータも、前述の検索子音響特徴パラメータと同様に、入力音響映像信号のうちから、使用者が指定した区間の映像信号を解析することにより得ることができるし、別途、検索子とする映像部分からなる映像信号を入力して、それを解析して得ることもできる。また、パーソナルコンピュータなどで予め解析して得られた検索子を、装置に登録するようにすることも、勿論できる。

そして、検索対象の映像信号を解析して得た映像特徴パラメータと、検索子映像特徴パラメータとを比較し、両者の類似性を算出することにより、検索対象の映像信号中から検索子の画像と類似する画像部分を検索するようにする。そして、音響信号についての検索

面は検索されたときにのみ表示しても良いし、常時表示しておいても良い)として、視聴している親画面の一部に表示するようにすることもできる。この機能は、例えば、お気に入りの俳優の声紋などを検索子音響パラメータとして用い、その俳優がでてくるシーンを子画面に表示するなどの使い方ができる。

また、図17に示した縮小静止画200のそれぞれに対応して、検索子音響特徴パラメータを登録しておき、使用者が検索子として縮小静止画200のいずれかを指示したとき、その指示された縮小静止画200に対応する検索子音響特徴パラメータを用いた音響検索動作を行うようにすることもできる。

例えば、縮小静止画には、特定の人物が表示されており、その人物の声紋などを検索子音響特徴パラメータとして検索子メモリに記憶しておき、その人物の縮小静止画が検索子として指定されたときに、音響検索処理を行って、当該人物が登場して声を出しているシーンを抽出することができる。このように、縮小静止画に対応して、検索音響特徴パラメータを登録しておくことにより、縮小静止画を検索のためのインデックスとして使用することができる。

また、逆に、特定の検索子音響特徴パラメータを指定することにより、その検索子音響特徴パラメータで検索されたシーンを、縮小静止画表示に使用する縮小静止画として登録するようにすることもできる。例えば、人物は映し出されていないが、特定の人物の声紋などの音響的特徴を含むシーンを、前記縮小静止画として登録することができる。

なお、検索子音響信号や検索子音響特徴パラメータを放送信号にデータとして重畳して提供するようにすることもできる。使用者は、

サ/デマルチプレクサ 100 に送出されて、デジタルインタフェース 74 でインタフェース処理されてデジタル入出力端子 73 に出力され、デジタル入出力端子 73 からシリアルバス 64 を介して外部装置 63 に送出されて、外部装置 63 に接続されたモニタ装置 7 において、画像が表示され、音声出力される。

これと同時に、マルチプレクサ/デマルチプレクサ 100 のデマルチプレクサで、再生されたトランスポートストリームから PES が分離されて、MP EG 映像音声デコーダ 91 に供給されることによって、出力端子 72 V, 72 S 及び 72 A に、再生アナログ映像信号及び再生アナログ音響信号を得ることができる。

(音響検索処理)

前述したように、音響検索処理部 150 により、前述した第 1 の実施の形態～第 5 の実施の形態と同様にして、音響検索することができる。この場合に、特定の検索子音響パラメータを指定して検索された部分の音響信号だけでなく、その検索された音響信号部分に対応する映像信号部分を、両者のリンク関係から、検索結果として抽出することができる。例えば、音楽番組の場合には、音楽が流れているシーン部分のみを検索して抽出することができる。

この場合、検索対象の音響信号は、ハードディスク装置 120 に記録された音響データだけではなく、受信したあるいは入力された音響データであっても良いことは、前述した通りである。この場合に、この実施の形態の場合には、複数の入力音響データが存在するので、ある入力音響映像信号による画像と音声とをモニタ装置 62 で視聴しているときに、他の入力音響映像信号中の音響信号について、検索を行い、その検索結果の画像部分を、例えば子画面（子画

ディスク 121～123 から再生されたトランスポートストリームを外部装置 63 に出力することができるように構成される。

外部装置 63 としては、IRD (Integrated Receiver Decoder) やパーソナルコンピュータなどの機器を接続することができる。デジタルインタフェース 74 としては、IEEE (Institute of Electrical and Electronics Engineers) 1394 デジタルインタフェースが用いられ、外部装置 63 とデジタル入出力端子 73 とを接続するシリアルバス 64 としては、IEEE 1394 シリアルバスが用いられる。

外部装置 63 からシリアルバス 64 を介してデジタル入出力端子 73 に入力されたトランスポートストリームは、デジタルインタフェース 74 でインタフェース処理されてマルチプレクサ/デマルチプレクサ 100 に送出され、マルチプレクサ/デマルチプレクサ 100 からバッファコントローラ 110 に送出されて、ハードディスク装置 120 によってハードディスク 121～123 に記録される。

これと同時に、マルチプレクサ/デマルチプレクサ 100 のデマルチプレクサで、記録されるトランスポートストリームから PES が分離されて、MP EG 映像音声デコーダ 91 に供給されることによって、出力端子 72V, 72S 及び 72A に、アナログ映像信号及びアナログ音響信号を得ることができる。

再生時には、ハードディスク装置 120 によってハードディスク 121～123 からトランスポートストリームが読み出されて、その再生されたトランスポートストリームが、バッファコントローラ 110 に送出され、バッファコントローラ 110 からマルチプレク

表画面の例を、図 17 に示す。この図 17 の例では、16 枚の縮小静止画 200 がウィンドウに貼り付けられて表示される。

NTSCエンコーダ 94 では、OSD 処理回路 93 からの映像データ（コンポーネント映像データ）が、輝度データと色差データとに変換された上で、又は変換されることなくそのまま、それぞれ D/A（Digital to Analog）変換されて、それぞれアナログ信号のセパレート映像信号及びコンポジット映像信号が生成される。そのセパレート映像信号は出力端子 72 S に導出され、コンポジット映像信号は出力端子 72 V に導出される。

一方、出力切換回路 95 では、システムコントローラ 140 によって、MP EG 映像音声デコーダ 91 からの音響データと、A/D 変換器 88 からの音響データとの、いずれかが選択されて取り出される。その選択された音響データは、D/A 変換器 36 でアナログ音響信号に変換されて、出力端子 72 A に導出される。

出力端子 72 S に導出されたセパレート映像信号、出力端子 72 V に導出されたコンポジット映像信号、及び出力端子 72 A に導出された音響信号が、外部のテレビジョン受信機などのモニタ装置 2 に出力されることによって、モニタ装置 2 において、画像が表示され、音声出力される。

#### （外部装置からの映像音響データの記録再生）

この例では、記録再生装置 70 は、デジタル入出力端子 73 とマルチプレクサ/デマルチプレクサ 100 との間にデジタルインタフェース 74 が接続されて、デジタル入出力端子 73 に外部装置 63 を接続することによって、外部装置 63 から入力されたトランスポートストリームをハードディスク 121～123 に記録し、ハード

プレクサ100からのPESが映像エレメンタリストリームと音声エレメンタリストリームに分離され、さらに、その映像エレメンタリストリームがMP EG映像デコーダで復号されてベースバンドの映像データに変換され、音声エレメンタリストリームがMP EG音声デコーダで復号されてベースバンドの音響データに変換される。変換後の映像データはポスト映像処理回路92に供給され、音響データは出力切換回路95に供給される。

ポスト映像処理回路92では、システムコントローラ140の指示によって、MP EG映像音声デコーダ91からの映像データ及びプリ映像処理回路86からの映像データに対して、両者の切り換え、又は合成や、ポストフィルタリング処理などの処理が施される。

ポスト映像処理回路92では、また、GUI (Graphical User Interface) 用に、代表画面 (静止画) を生成し、複数の代表画面を縮小して、ウィンドウに貼り付ける縮小静止画表示処理が施され、処理後の映像データが、OSD (On Screen Display) 処理回路93に供給される。

代表画面としては、例えば番組タイトルの中から特徴的な1場面や、番組の中の見たい場면을素早くアクセスできるようにするための、その番組のインデックスポイント (ワンシーン) の画面等がある。

OSD処理回路93では、システムコントローラ140の指示によって、テキスト情報などに対応する画像データを生成して、ポスト映像処理回路92からの映像データの、上記のウィンドウ部分の映像データに重畳する処理が施され、処理後の映像データが、NTSCエンコーダ34に供給される。縮小静止画表示された複数の代

そのトランスポートストリームを入力レートより高いレートでバッファから読み出してハードディスク装置120に送出する。これによって、バッファコントローラ110に連続的に入力されるトランスポートストリームは、途切れることなくハードディスク121～123に記録される。

ハードディスク装置120は、システムコントローラ140によって制御されて、ハードディスク121～123にトランスポートストリームを書き込む。バッファコントローラ110とハードディスク装置120との間のプロトコル（インタフェース）としては、IDE（Integrated Drive Electronics）などが用いられる。

#### （再生及びアナログ出力処理）

再生時には、ハードディスク装置120は、システムコントローラ140によって制御されて、ハードディスク121～123からトランスポートストリームを読み出し、バッファコントローラ110に送出する。バッファコントローラ110は、記録時とは逆に、ハードディスク装置120から断続的に入力されるトランスポートストリームを、連続的なトランスポートストリームに変換して、マルチプレクサ／デマルチプレクサ100に供給する。

マルチプレクサ／デマルチプレクサ100では、そのデマルチプレクサにおいて、その連続的なトランスポートストリームのヘッダが解析処理されることによって、トランスポートストリームからPES（Packetized Elementary Stream）が分離されて、MPEG映像音声デコーダ91に供給される。

MPEG映像音声デコーダ91では、マルチプレクサ／デマルチ

MPEG映像エンコーダ87では、プリ映像処理回路86からの映像データに対してブロックDCT (Discrete Cosine Transform) などの符号化処理が施されて、映像エレメンタリストリーム (ES: Elementary Stream) が生成され、その映像エレメンタリストリームがマルチプレクサ/デマルチプレクサ100に供給される。

一方、入力切換回路82からの音響信号は、A/D変換器88でデジタル音響データに変換されたのち、MPEG音声エンコーダ89及び出力処理部90の出力切換回路95に供給される。

MPEG音声エンコーダ89では、A/D変換器88からの音響データがMPEG方式によって圧縮符号化されて、音声エレメンタリストリームが生成され、その音声エレメンタリストリームがマルチプレクサ/デマルチプレクサ100に供給される。

マルチプレクサ/デマルチプレクサ100では、そのマルチプレクサにおいて、MPEG映像エンコーダ87からの映像エレメンタリストリーム及びMPEG音声エンコーダ89からの音声エレメンタリストリームと各種の制御信号が多重化されて、MPEGシステムのトランスポートストリームが生成され、そのトランスポートストリームがバッファコントローラ110に送出される。

バッファコントローラ110は、マルチプレクサ/デマルチプレクサ100から連続的に入力されるトランスポートストリームを、ハードディスク装置120に断続的に送出する。すなわち、ハードディスク装置120がシーク動作を行っているときには、書き込みができないので、入力されたトランスポートストリームをバッファに一時蓄え、ハードディスク装置120が書き込み可能なときに、

入力切換回路 8 2 からのコンポジット映像信号は、Y C 分離回路 8 3 で輝度信号と色差信号に分離されて、別の入力切換回路 8 4 の一方の入力端に供給される。また、入力端子 7 1 S からのセパレート映像信号（輝度信号及び色差信号）が、入力切換回路 8 4 の他方の入力端に供給され、入力切換回路 8 4 がシステムコントローラ 1 4 0 によって切り換えられて、入力切換回路 8 4 から、いずれかの輝度信号及び色差信号が選択されて取り出される。

入力切換回路 8 4 からの輝度信号及び色差信号は、NTSC (National Television System Committee) デコーダ 8 5 で、それぞれ A/D (Analog to Digital) 変換され、さらにクロマエンコード処理されて、NTSC デコーダ 8 5 からコンポーネント映像データが得られる。

また、NTSC デコーダ 8 5 では、入力切換回路 8 4 からの輝度信号から垂直同期信号及び水平同期信号が分離されるとともに、その分離された同期信号に基づいてクロック及びフィールド判別信号が生成され、これら同期信号、クロック及びフィールド判別信号が、同期制御回路 1 3 0 に供給される。同期制御回路 1 3 0 では、これらの信号を基準として記録再生装置 7 0 の各部に必要なクロック及びタイミング信号が生成されて、記録再生装置 7 0 の各部に供給される。

NTSC デコーダ 8 5 からの映像データは、プリ映像処理回路 8 6 でプリフィルタリング処理などの処理が施されたのち、MP EG 映像エンコーダ 8 7 及び出力処理部 9 0 のポスト映像処理回路 9 2 に供給される。

索処理を行う部分により、この音響検索処理部 150 が構成される。

なお、検索子メモリ 32 は、音響検索子処理部 150 内に設けることもできるし、ハードディスク装置 120 のハードディスクの一部の記憶エリアを用いることもできる。

なお、デジタル音響圧縮データや、音響データと映像データとがリンクされて記録再生される音響映像データでは、検索や登録の該当区間を示すリンクされたアドレス情報として、フレーム番号やフィールド番号、PTS (Presentation Time Stamp)、パケット番号などが用いられる。

(アナログ入力処理及び記録)

アンテナ 61 で受信されたアナログテレビジョン放送の信号は、操作入力部 160 での操作に基づくシステムコントローラ 140 の指示によって、チューナ 81 で選局されて、チューナ 81 から、選択されたチャンネルのコンポジット映像信号及び音響信号が得られ、そのコンポジット映像信号及び音響信号が、入力切換回路 82 の一方の入力端に供給される。

入力端子 71 V には、外部機器からコンポジット映像信号が、入力端子 71 A には、外部機器から音響信号が、入力端子 71 S には、外部機器からセパレート映像信号 (輝度信号と色差信号とに分離されたもの) が、それぞれ供給される。

入力端子 71 V からのコンポジット映像信号、及び入力端子 71 A からの音響信号は、入力切換回路 82 の他方の入力端に供給され、入力切換回路 82 がシステムコントローラ 140 によって切り換えられて、入力切換回路 82 から、いずれかのコンポジット映像信号及び音響信号が選択されて取り出される。

再生されるように構成される。

システムコントローラ140は、CPU (Central Processing Unit) 141を備え、そのシステムバス142にROM (Read Only Memory) 143及びRAM (Random Access Memory) 144などが接続されて構成される。

ROM143には、CPU141が実行すべきプログラムが、あらかじめ格納され、CPU141は、そのプログラムを読み出して、記録再生装置70の各部を制御する。RAM144には、CPU141が記録再生装置70の各部を制御する上で必要な、その他のプログラムやデータが適宜、書き込まれる。

操作入力部160は、キーボード、マウス、ボタン、スイッチ、リモートコマンドなどによって構成され、図示を省略した入出力インタフェースを介してシステムコントローラ140のシステムバス142に接続される。

音響検索処理部150は、前述した第1の実施の形態から第5の実施の形態のうちのいずれかの機能を備えるものとして構成される部分である。特に、この実施の形態はMP EG規格の圧縮符号化データが検索対象であるので、前述したように、第3の実施の形態の場合が好適である。

ただし、この場合、記録媒体13は、ハードディスク装置120のハードディスクを用いるようにされており、このため、記憶部12又は42、再生部14又は44などはバッファコントローラ110の部分により構成されるので、前述の各実施の形態の図のうち、主データとしての音響データの記録再生に関する部分を除く音響検

の入力信号とすることもできる。

図 1 6 は、M P E G 2 規格で圧縮された音響映像信号を記録再生する記録再生装置に、この実施の形態の音響映像検索装置及び音響映像蓄積装置を適用した場合の構成例を示すブロック図である。

この図 1 に示した記録再生装置は、映像信号及び音響信号を、M P E G 2 規格によって圧縮符号化し、多重化して、M P E G システムのトランスポートストリーム (T S : T r a n s p o r t S t r e a m) として、ハードディスク装置 (H D D : H a r d D i s k D r i v e) 内のハードディスクに記録し、ハードディスクから再生する場合である。

なお、以下では、アナログ映像信号及びアナログ音響信号を、映像信号及び音響信号と称し、デジタル映像信号及びデジタル音響信号は、デジタル映像データ及びデジタル音響データ、又は単に映像データ及び音響データと称する。

この例では、記録再生装置 7 0 は、アナログテレビジョン放送を受信するアンテナ 6 1 が接続され、アナログ入力端子 7 1 V, 7 1 A, 7 1 S、アナログ出力端子 7 2 V, 7 2 A, 7 2 S、デジタル入出力端子 7 3、デジタルインタフェース 7 4、入力処理部 8 0、出力処理部 9 0、マルチプレクサ/デマルチプレクサ 1 0 0、バッファコントローラ 1 1 0、ハードディスク装置 1 2 0、同期制御回路 1 3 0、システムコントローラ 1 4 0、音響検索処理部 1 5 0、操作入力部 1 6 0 を備える。

ハードディスク装置 1 2 0 は、この例では、3 枚のハードディスク 1 2 1, 1 2 2, 1 2 3 を備え、そのいずれかにトランスポートストリームが記録され、いずれかからトランスポートストリームが

また、上述の説明では、検索結果を記録媒体に記録するようにしたが、コントロール部が管理するメモリに、記録媒体に対応させて、検索結果を記録するようにするようによい。

#### [その他の実施の形態]

以上の実施の形態では、検索子は、予めメモリに登録したり、別個に用意するように説明したが、例えば、音響信号の記録中に、使用者が、その音響信号について検索子の音響信号部分を指定したとき、その指定後の音響信号については、指定した検索子を用いて検索を行うようにすることもできる。その場合には、予め検索子に登録しておく必要はない。

また、検索子メモリに登録する検索子は、入力音響信号や再生音響信号から抽出する必要もない。例えば、検索子としたい音源を用意し、その音源からの音響信号を用いて、検索子に登録することもできる。

また、検索子として登録する音響信号や音響特徴パラメータを、例えば、インターネットなどのネットワークを通じて取得して、検索子メモリに登録するようにすることもできる。

また、例えば記録済み記録媒体を販売するときに、その記録媒体に、検索子の候補を予め一緒に記録するようによい。また、その記録媒体を販売するときに、検索子の候補を記録したメモリカードなどを付加して販売するようによい。

#### [音響映像検索装置、音響映像蓄積装置の第1の実施の形態]

以上説明した実施の形態では、検索対象の入力信号は、音響信号単独の場合であったが、音響信号が、映像信号に関連するものとしてリンクされて、例えば記録再生される音響映像信号を、検索対象

ときに開始となる。

まず、記録媒体から検索用インデックステーブルを作成するための検索用インデックスのデータが読み出される（ステップS 3 0 1）。コントロール部 1 0 は、読み出したデータを基に、図 1 4 に示したような検索用インデックステーブルを作成し、表示画面に表示する（ステップS 3 0 2）。

使用者は、表示された検索用インデックステーブルを見て、検索再生したい検索子をキー操作部を通じて入力する。そこで、使用者により指定された検索子を認識する（ステップS 3 0 3）。そして、検索用インデックステーブルを参照し、その認識した検索子について、類似区間として判定された区間データを認識する（ステップS 3 0 4）。そして、その認識した各区間データで示されるアドレス区間の音響信号を、記録媒体から順次読み出して、再生する（ステップS 3 0 5）。

こうして、この第 5 の実施の形態によれば、検索子に類似する音響信号部分のみを簡単に抽出して再生することができる。もっとも、この第 5 の実施の形態の場合にも、検索用インデックステーブルにない検索子を検索子メモリから指定して、検索を行うことも勿論可能である。

なお、上述の第 5 の実施の形態の説明では、検索子の識別情報と検索子名とを検索用インデックステーブルに含めるようにしたが、検索子の音響信号を検索子名に替えて、あるいは併せて記録するようにしてもよい。そのようにすれば、検索子がどのような音響信号であったかを、それを音響再生することにより、容易に知ることができる。

**This Page Blank (uspto)**

5/15

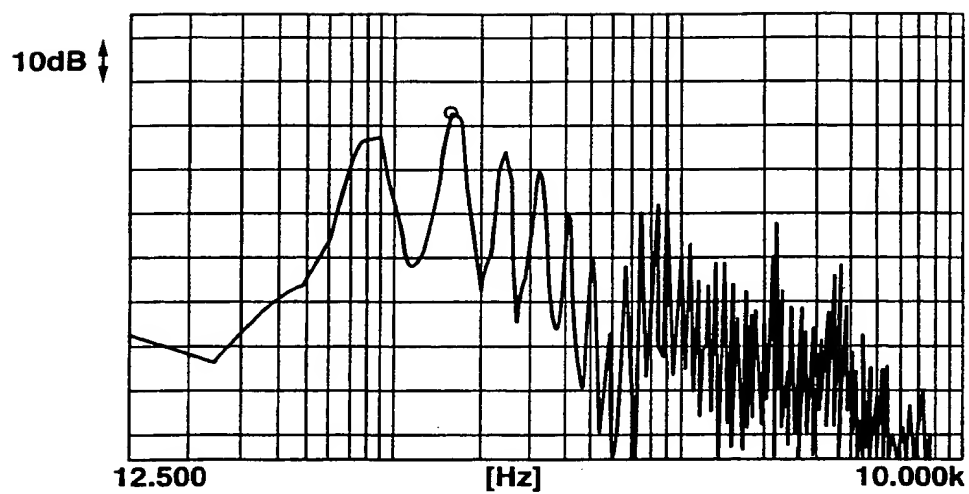


FIG.5A

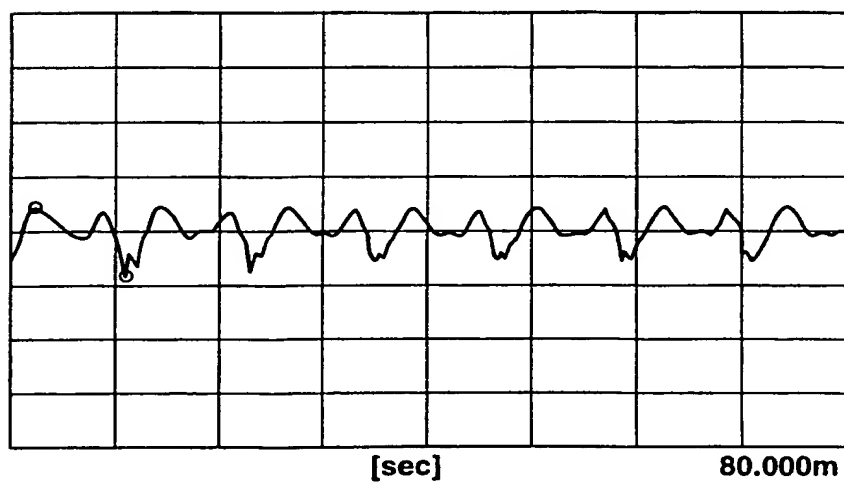
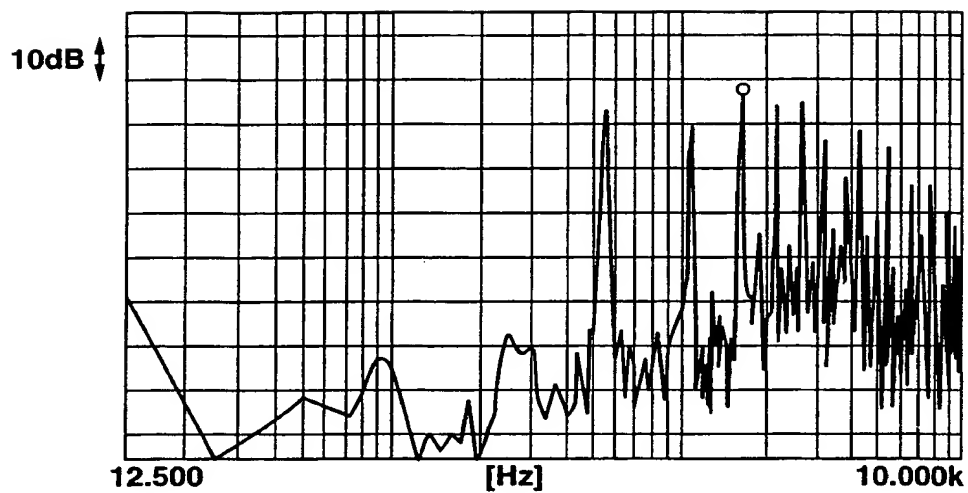
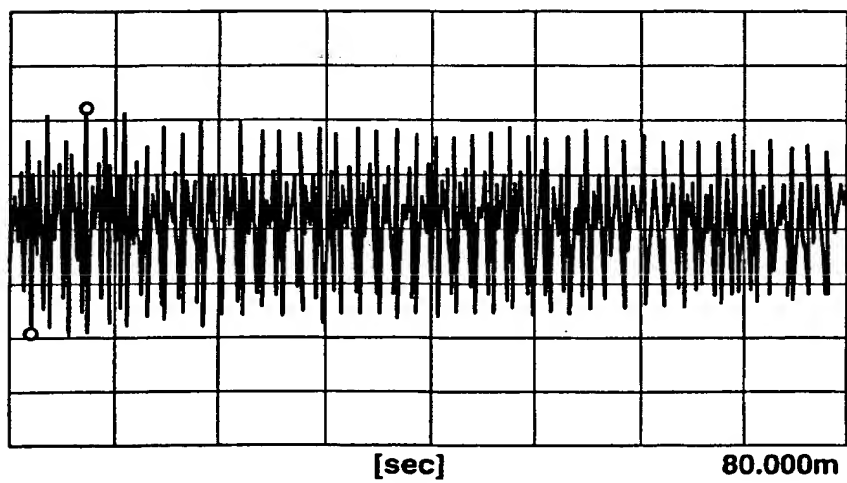


FIG.5B

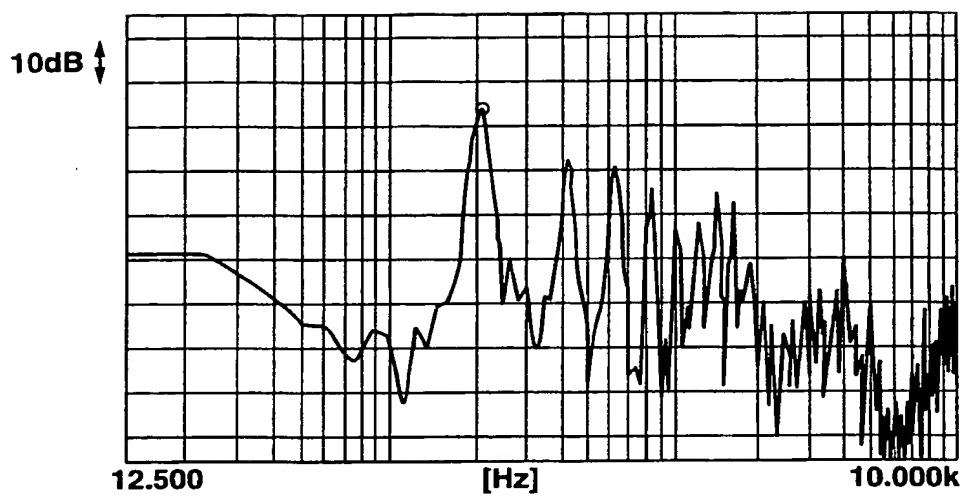
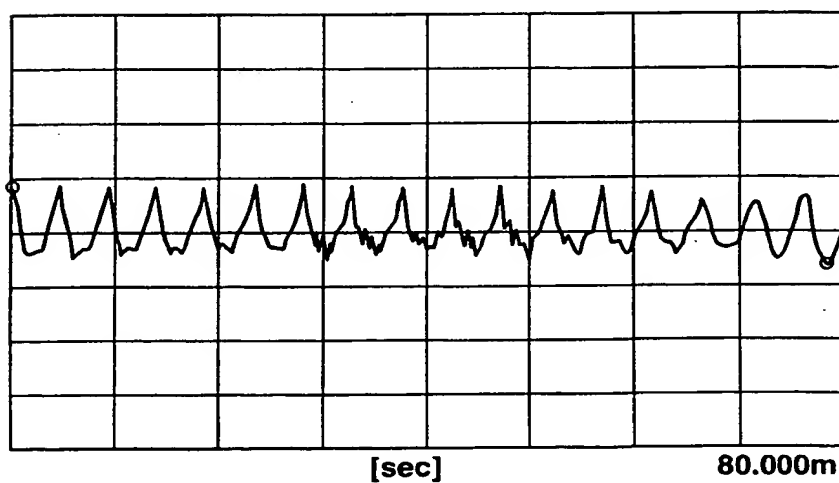
**This Page Blank (us<sub>7</sub>**

4/15

**FIG.4A****FIG.4B**

**This Page Blank (usp**

3/15

**FIG.3A****FIG.3B**

**This Page Blank (uspto)**

2/15

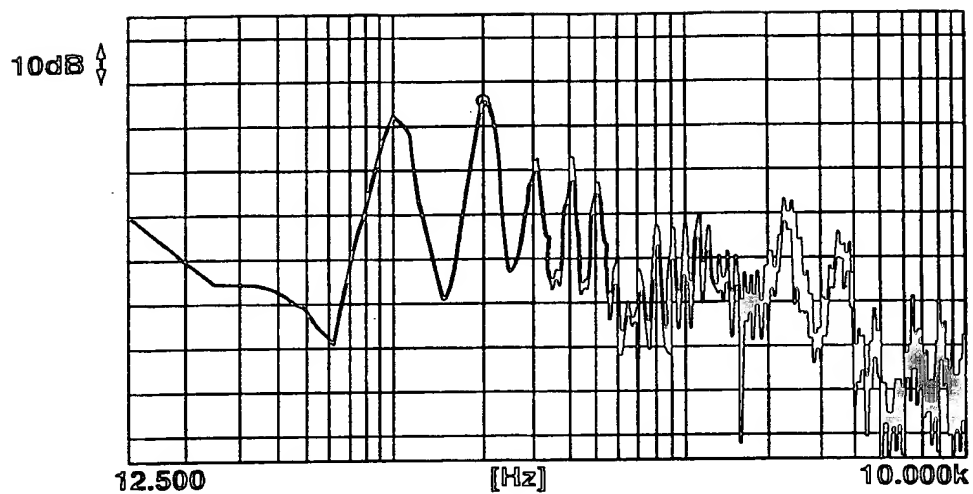


FIG.2A

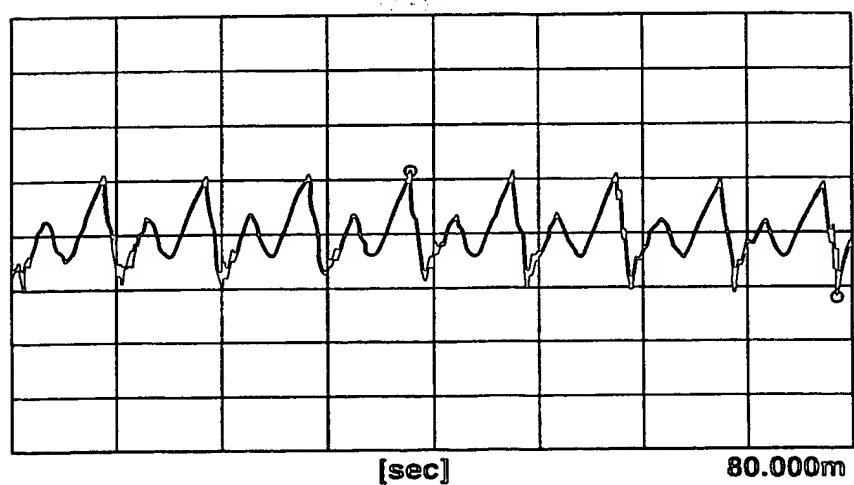
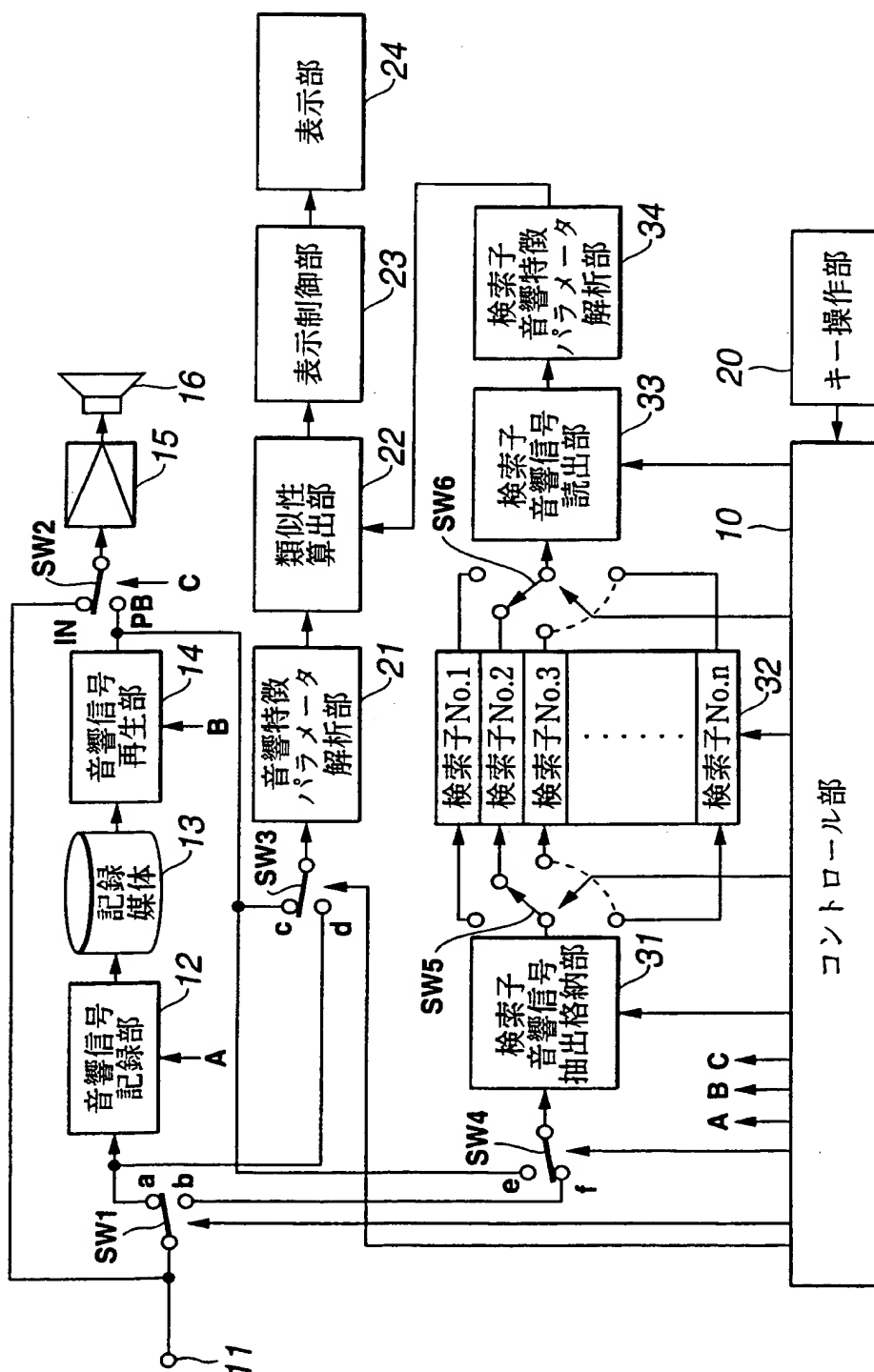


FIG.2B

**This Page Blank (uspto)**



**FIG. 1**

数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力信号中の音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索する音響部分検索手段とを備え、前記音響部分検索手段の検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする音響映像情報蓄積装置。

47. 映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録する記録手段と、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索する音響部分検索手段と、検索子とする1～複数の映像信号についての、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、検索対象の前記入力信号中の映像信号を解析して得た前記映像特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の映像信号中から前記検索子の画像と類似する画像部分を検索する映像部分検索手段とを備え、前記音響部分検索手段及び映像部分検索手段の検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする音響映像情報蓄積装置。

像信号の映像特徴パラメータとを比較して、前記入力信号中の映像信号中から前記検索子の映像信号と類似する部分を検索する映像部分検索手段と、前記音響部分検索手段及び映像部分検索手段の結果を表示する結果表示手段とを備える音響映像情報検索装置。

45. 映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音源についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記入力信号中の音響信号の音響特徴パラメータとを比較して、前記入力信号中の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する音響部分検索手段と、前記入力信号中の映像信号を、画像の特徴を表す映像特徴パラメータに解析する映像特徴パラメータ解析手段と、検索子とする1～複数の映像信号についての前記映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、前記映像特徴パラメータ解析手段からの前記入力信号中の映像信号の映像特徴パラメータとを比較して、前記入力信号中の映像信号中から前記検索子の映像信号と類似する部分を検索する映像部分検索手段と、前記音響部分検索手段及び映像部分検索手段での検索結果に基づいて決定される前記入力信号部分を抽出して再生する再生手段とを備える音響映像情報検索装置。

46. 映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録する記録手段と、前記入力信号中の音響信号を解析して、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする1～複

4 2. 前記結果表示手段は、前記検索手段で検索された部分の映像信号による画像を表示することを特徴とする請求の範囲第 4 1 項記載の音響映像情報検索装置。

4 3. 映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段での検索結果に基づいて、前記類似する部分に対応する前記入力信号部分を抽出して再生する再生手段とを備える音響映像情報検索装置。

4 4. 映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音源についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記入力信号中の音響信号の音響特徴パラメータとを比較して、前記入力信号中の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する音響部分検索手段と、前記入力信号中の映像信号を、画像の特徴を表す映像特徴パラメータに解析する映像特徴パラメータ解析手段と、検索子とする 1 ～複数の映像信号についての前記映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、前記映像特徴パラメータ解析手段からの前記入力信号中の映

40. 映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録するとともに、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索し、さらに、検索子とする1～複数の映像信号についての、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、検索対象の前記入力信号中の映像信号を解析して得た前記映像特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の映像信号中から前記検索子の画像と類似する画像部分を検索し、前記音響信号及び映像信号についての検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする音響映像情報蓄積方法。

41. 映像信号と、その映像信号に関連する音響信号とからなる入力信号中の前記音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段の結果を表示する結果表示手段とを備える音響映像情報検索装置。

て、検索子とする 1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索するとともに、検索子とする 1～複数の映像信号についての、画像の特徴を表す映像特徴パラメータからなる検索子映像特徴パラメータのそれぞれと、検索対象の前記入力信号中の映像信号を解析して得た前記映像特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の映像信号中から前記検索子の画像と類似する画像部分を検索し、前記音響信号についての検索結果と、前記映像信号についての検索結果とに基づいて、前記入力信号中から所定の部分を検索することを特徴とする音響映像検索処理方法。

39. 映像信号と、その映像信号に関連する音響信号とからなる入力信号を記録媒体に記録するとともに、前記入力信号中の音響信号を解析して、音源の構造的な要素に依存する音響特徴パラメータを得、検索子とする 1～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力信号中の音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索し、その検索結果を、前記検索子と前記入力信号の前記記録媒体上の記録位置とに関連させて記録するようにすることを特徴とする音響映像情報蓄積方法。

子の音響信号から検索子音響特徴パラメータを得る検索子音響特徴パラメータ解析手段とを備え、前記検索子音響特徴パラメータ解析手段で求められた検索子音響特徴パラメータを前記検索手段で前記入力音響信号についての前記音響特徴パラメータと比較することを特徴とする請求の範囲第31項又は第32項のいずれか1項に記載の音響情報蓄積装置。

35. 前記検索子の音響信号を、前記検索対象の音響信号から、使用者の指示に従って抽出して前記検索子音響信号保持部に保持する検索子登録手段を設けたことを特徴とする請求の範囲第34項に記載の音響情報蓄積装置。

36. 前記検索子は複数個である場合に、予め設定された単位時間ごとに、検索子を切り換えて、前記類似性の算出を行うことを特徴とする請求の範囲第31項又は第32項のいずれか1項に記載の音響情報蓄積装置。

37. 映像信号と、その映像信号に関連する音響信号とからなる入力信号から、所定の部分を検索する音響映像検索処理方法において、検索子とする1～複数の音響信号についての、音源の構造的な要素に依存する音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、検索対象の前記入力信号中の音響信号を解析して得た前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記入力信号中の音響信号中から前記検索子の音響信号と類似する部分を検索し、その検索した部分を、前記入力信号についての検索結果とすることを特徴とする音響映像検索処理方法。

38. 映像信号と、その映像信号に関連する音響信号とからなる入力信号から、所定の部分を検索する音響映像検索処理方法におい

音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする1～複数の音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータのそれぞれと、前記入力音響信号についての前記音響特徴パラメータとを比較し、両者の類似性を算出することにより、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段とを備え、前記検索手段の検索結果を、前記検索子及び前記入力音響信号の前記記録媒体上の記録位置に関連させて記録するようにすることを特徴とする音響情報蓄積装置。

32. 前記検索結果は、前記入力音響信号とともに、前記記録媒体に記録することを特徴とする請求の範囲第31項に記載の音響情報蓄積装置。

33. 使用者の指示に応じて前記入力音響信号から選択抽出した音響信号部分についての解析を行って、前記検索子とする1～複数の音響信号についての前記検索子音響特徴パラメータのそれぞれを得る検索子音響特徴パラメータ解析手段と、前記検索子音響特徴パラメータ解析手段で求められた検索子音響特徴パラメータを、前記検索子として使用するために保持する検索子音響特徴パラメータ保持手段とを備えることを特徴とする請求の範囲第31項又は第32項のいずれか1項に記載の音響情報蓄積装置。

34. 前記入力音響信号から、使用者の指示に応じて選択抽出した音響信号部分を、前記検索子として抽出する検索子抽出手段と、前記検索子抽出手段で抽出された検索子の音響信号を記憶する検索子音響信号保持手段と、前記検索子保持手段から読み出された検索

いて、前記類似する部分のみを抽出して再生する再生手段とを備える音響情報検索装置。

26. 1～複数の前記検索子とする音響信号を保持する保持部と、検索時に、前記保持部から選択された検索子の音響信号を読み出す読み出し手段と、前記読み出し手段により読み出された検索子の音響信号を解析して、前記類似性算出手段に供給する検索子音響特徴パラメータを得る検索子音響特徴パラメータ解析手段とを備えることを特徴とする請求の範囲第25項に記載の音響情報検索装置。

27. 前記検索子の音響信号を、前記検索対象の音響信号から、使用者の指示に従って抽出して前記保持部に保持する検索子登録手段を設けたことを特徴とする請求の範囲第26項に記載の音響情報検索装置。

28. 1～複数の前記検索子音響特徴パラメータを保持する保持部と、検索時に、前記保持部から選択された検索子音響特徴パラメータを読み出し、前記類似性算出手段に供給する読み出し手段とを備えることを特徴とする請求の範囲第25項に記載の音響情報検索装置。

29. 前記検索子の音響特徴パラメータを、前記検索対象の音響信号から使用者の指示に従って抽出し、解析して得て、前記保持部に保持する検索子登録手段を設けたことを特徴とする請求の範囲第28項に記載の音響情報検索装置。

30. 前記検索子は複数個である場合に、予め設定された単位時間ごとに、検索子を切り換えて、前記類似性の算出を行うことを特徴とする請求の範囲第25項に記載の音響情報検索装置。

31. 入力音響信号を記録媒体に記録する記録手段と、前記入力

21. 1～複数の前記検索子とする音響信号を保持する保持部と、検索時に、前記保持部から選択された検索子の音響信号を読み出す読み出し手段と、前記読み出し手段により読み出された検索子の音響信号を解析して、前記類似性算出手段に供給する検索子音響特徴パラメータを得る検索子音響特徴パラメータ解析手段とを備えることを特徴とする請求の範囲第19項に記載の音響情報検索装置。

22. 1～複数の前記検索子音響特徴パラメータを保持する保持部と、検索時に、前記保持部から選択された検索子音響特徴パラメータを読み出し、前記類似性算出手段に供給する読み出し手段とを備えることを特徴とする請求の範囲第19項に記載の音響情報検索装置。

23. 前記検索子の音響特徴パラメータを、前記検索対象の音響信号から使用者の指示に従って抽出し、解析して得て、前記保持部に保持する検索子登録手段を設けたことを特徴とする請求の範囲第22項に記載の音響情報検索装置。

24. 前記検索子は複数個である場合に、予め設定された単位時間ごとに、検索子を切り換えて、前記類似性の算出を行うことを特徴とする請求の範囲第19項に記載の音響情報検索装置。

25. 検索対象の音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段での検索結果に基づ

メモリから、前記検索子とする 1～複数の音響信号がそれぞれ読み出されるとともに、それぞれの音響信号が解析されて検索子音響特徴パラメータとされ、それらの検索子音響特徴パラメータを用いた前記入力音響信号についての前記類似性の算出処理による前記類似する部分の検索処理が行われることを特徴とする請求の範囲第 14 項に記載の音響情報蓄積方法。

17. 前記検索結果とともに、前記検索子となる音響信号の音響特徴パラメータあるいは前記検索子となる音響信号部分も、前記記録媒体に記録することを特徴とする請求の範囲第 12 項又は請求項第 13 項のいずれか 1 項に記載の音響情報蓄積方法。

18. 前記検索結果とともに、前記検索子となる音響信号の音響特徴パラメータあるいは前記検索子となる音響信号部分も、前記記録媒体に記録することを特徴とする請求の範囲第 14 項に記載の音響情報蓄積方法。

19. 検索対象の音響信号を、音源の構造的な要素に依存する音響特徴パラメータに解析する音響特徴パラメータ解析手段と、検索子とする音響信号についての前記音響特徴パラメータからなる検索子音響特徴パラメータと、前記音響特徴パラメータ解析手段からの前記検索対象の音響信号の音響特徴パラメータとを比較して、前記検索対象の音響信号中から前記検索子の音源からの音響信号と類似する部分を検索する検索手段と、前記検索手段の結果を表示する結果表示手段とを備える音響情報検索装置。

20. 前記結果表示手段は、前記検索手段で算出された前記類似する部分を、前記検索対象の音響信号の時間の流れに応じて表示することを特徴とする請求の範囲第 19 項に記載の音響情報検索装置。

**This Page Blank (uspto)**

15/15

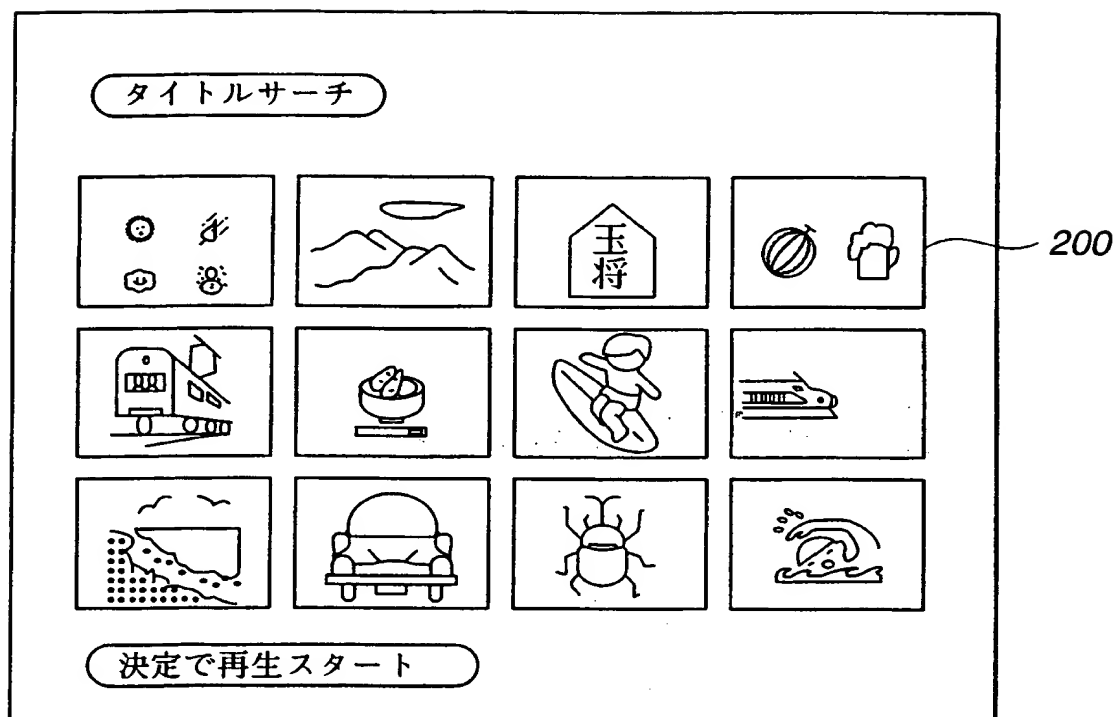


FIG.17

**This Page Blank (uspto)**

14/15

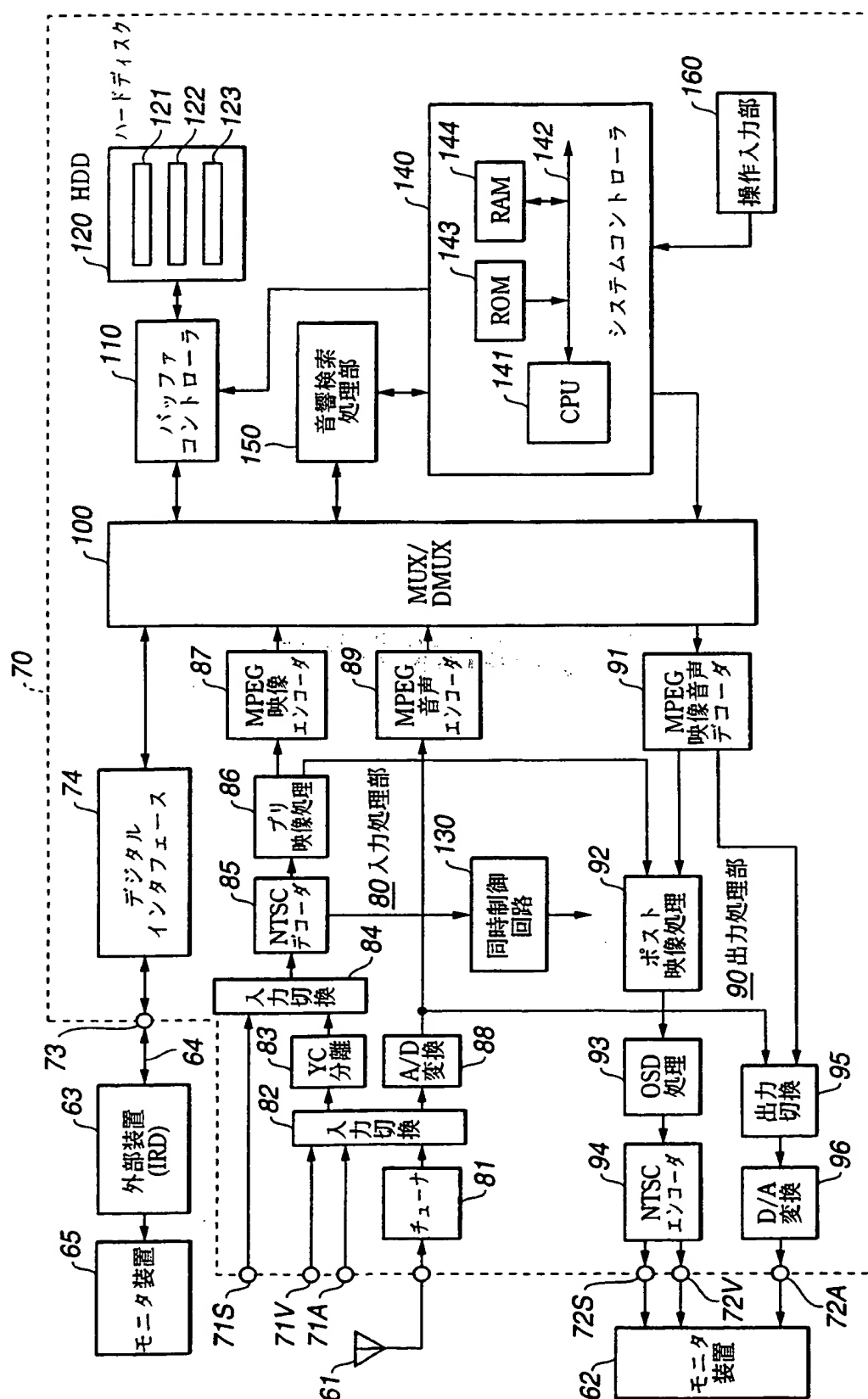


FIG.16

**This Page Blank (uspto)**

13/15

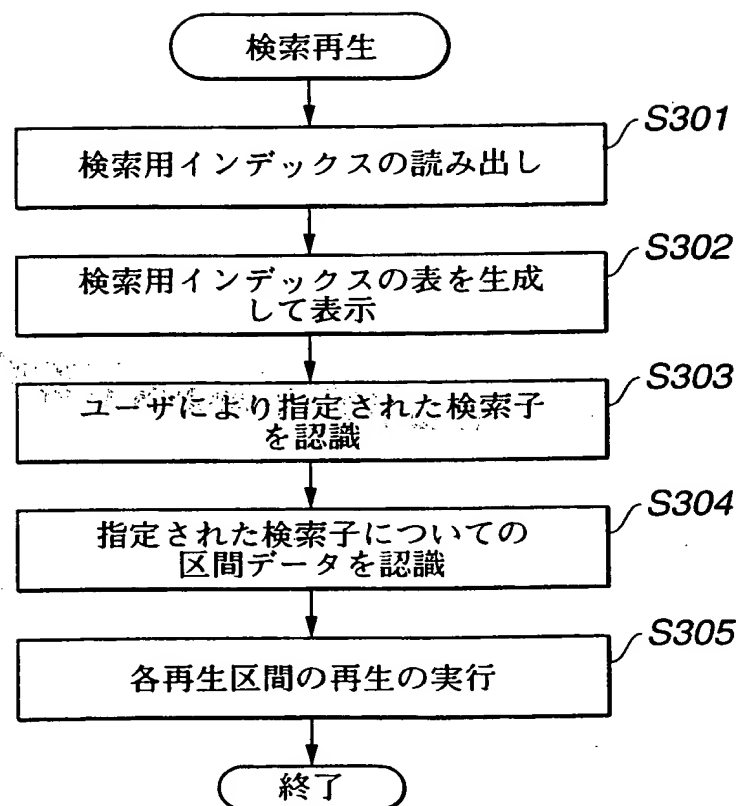


FIG.15

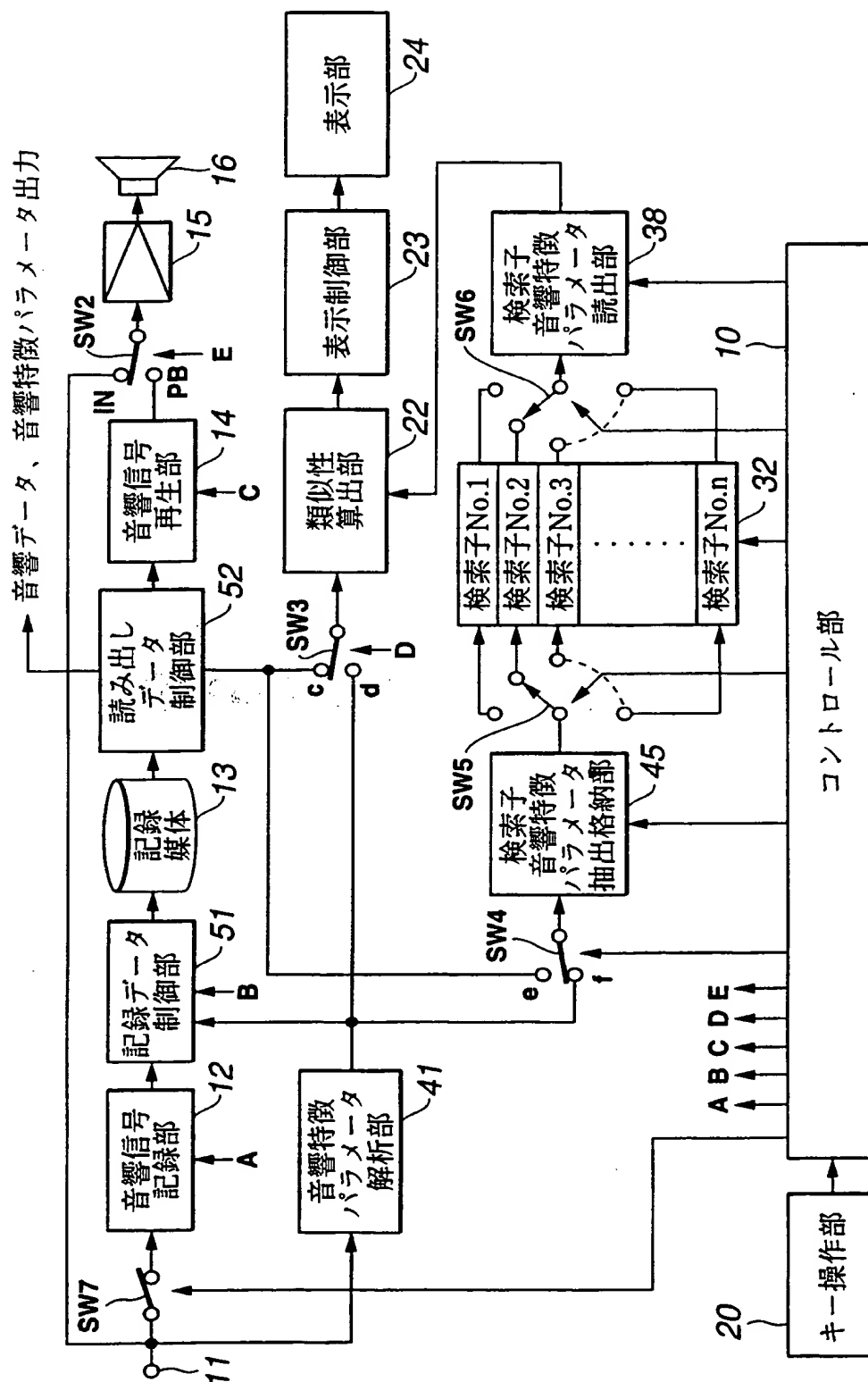
**This Page Blank (uspto)**

12/15

検索子No.	検索子名	区間データ[ ST , ED ]
1	S	[ ADR1 , ADR5 ] , [ ADR11 , ADR20 ]
2	T	[ ADR1 , ADR10 ] , [ ADR41 , ADR45 ]
3	U	[ ADR23 , ADR29 ]
⋮	⋮	⋮
n	Z	[ ADR101 , ADR108 ]

FIG.14

**This Page Blank (uspto)**



**FIG. 13**

**This Page Blank (uspto)**

10/15

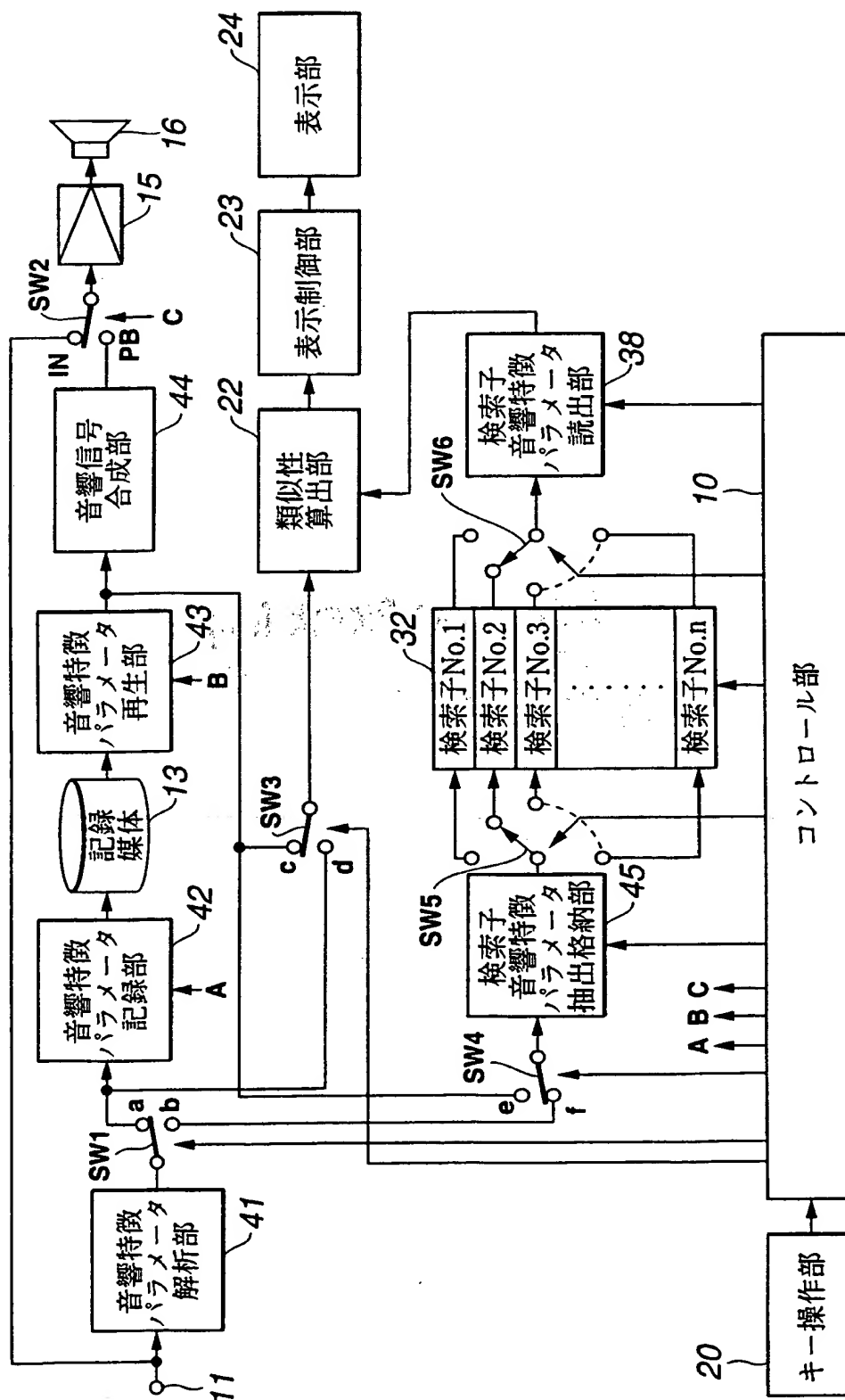


FIG.12

**This Page Blank (uspto)**

9/15

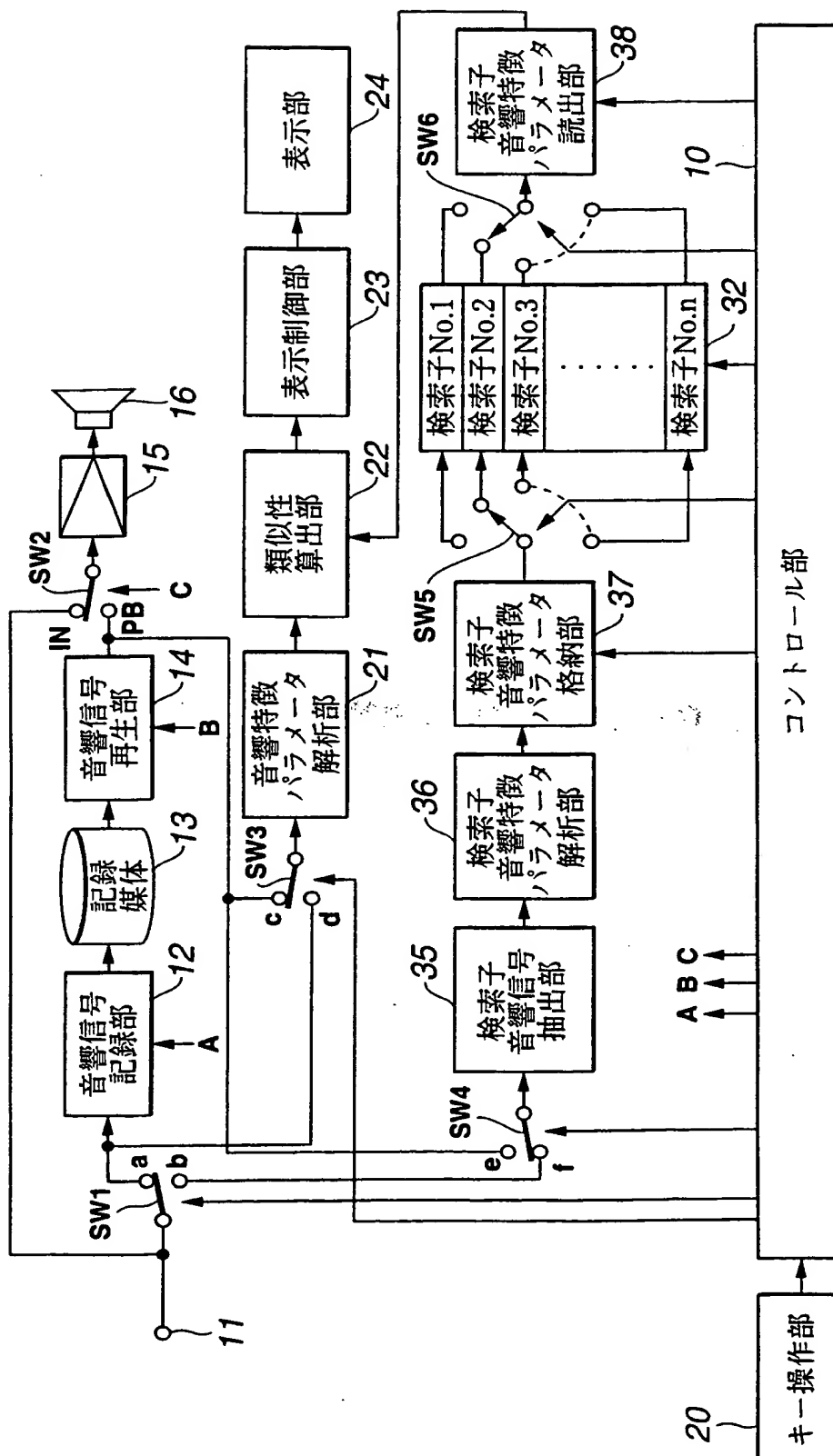


FIG.11

**This Page Blank (uspto)**

FIG.8A

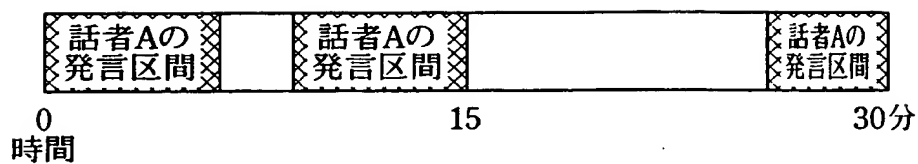


FIG.8B

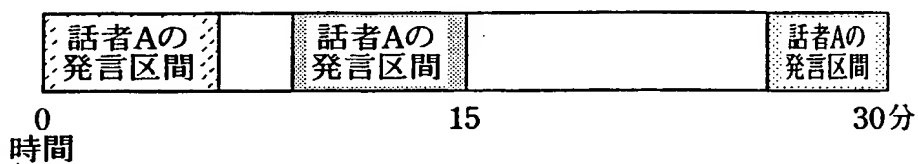


FIG.9

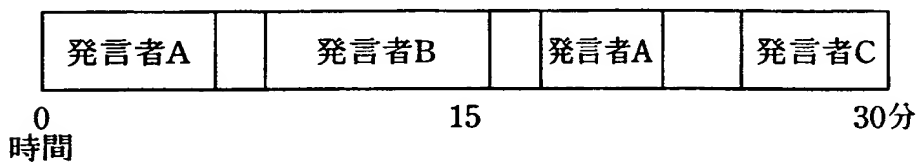
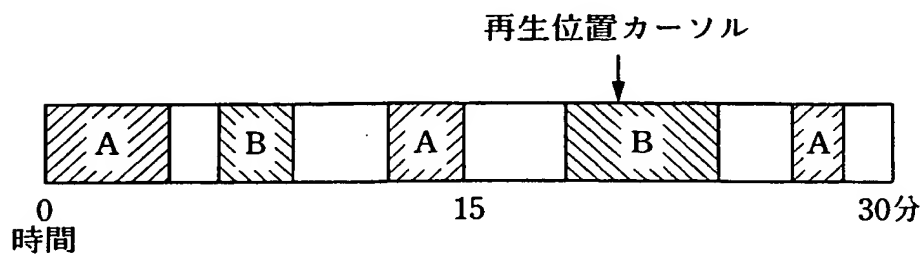


FIG.10



**This Page Blank (uspto)**

7/15

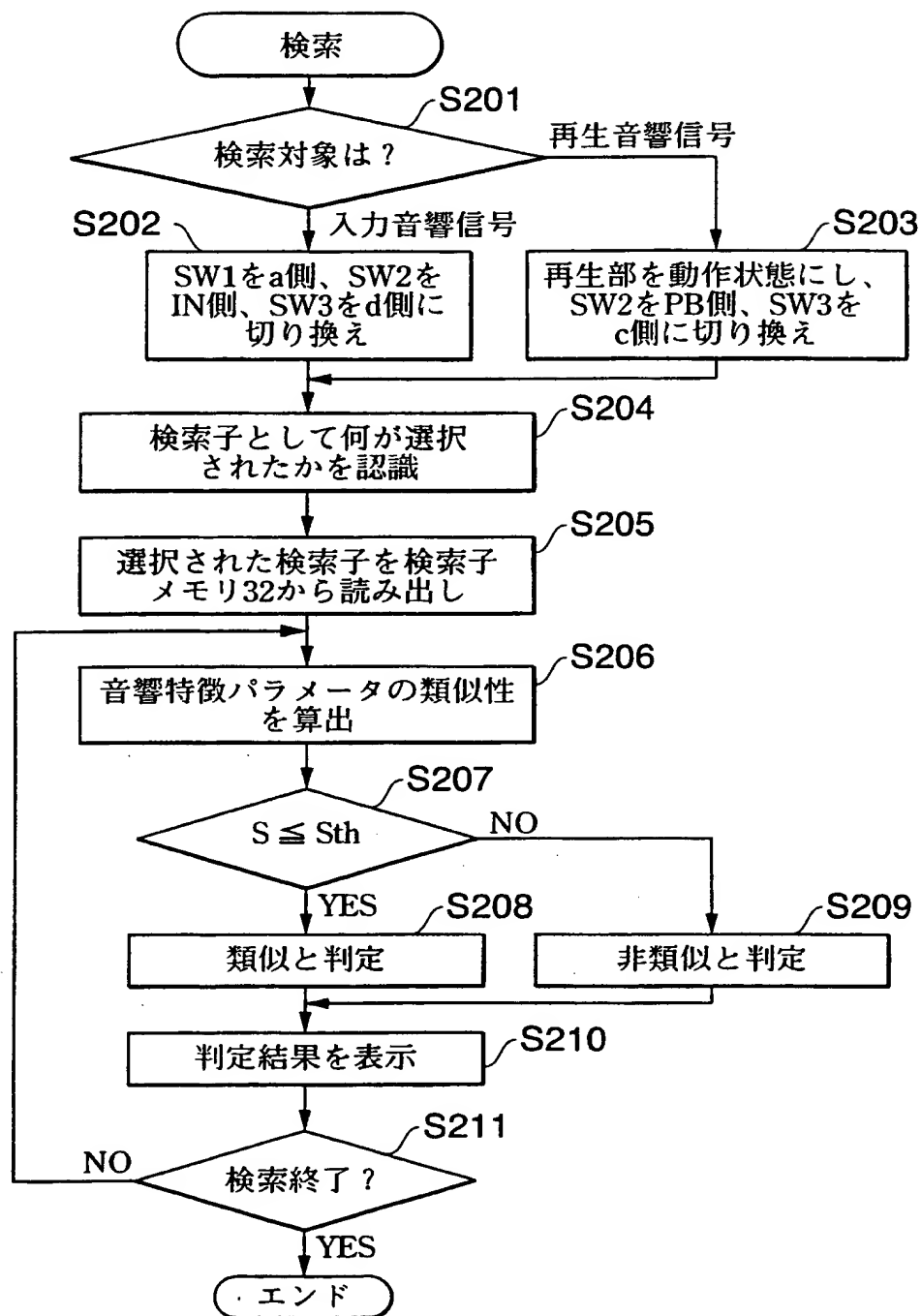


FIG.7

**This Page Blank (uspto)**

6/15

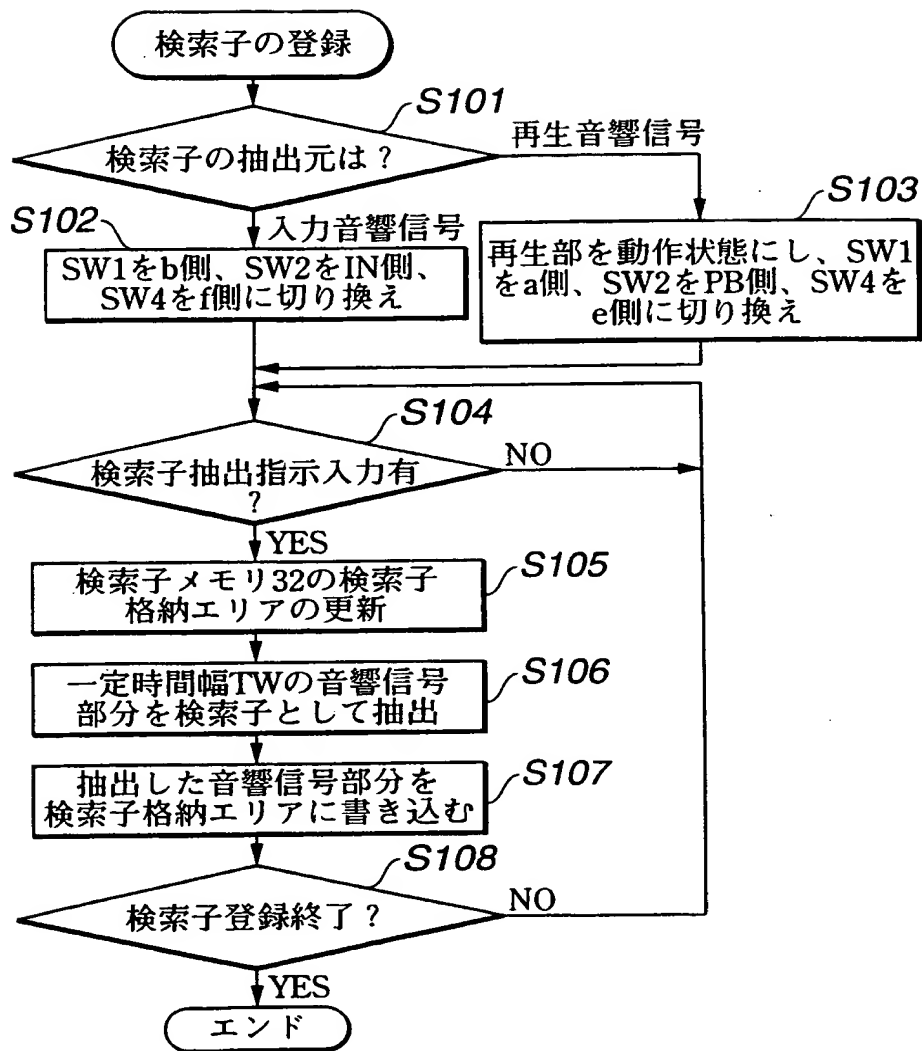


FIG.6

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/05771

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl<sup>7</sup> G10L15/00, G11B27/30, H04N5/76, H04N5/91

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> G10L15/00~17/00, G11B27/10~27/32,  
H04N5/76, H04N5/91

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1926-1995 Toroku Jitsuyo Shinan Koho 1994-2000  
Kokai Jitsuyo Shinan Koho 1971-2000 Jitsuyo Shinan Toroku Koho 1996-2000

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS) INSPEC (DIALOG)

WPI (DIALOG)

IEEE/IEE Electronic Library

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Proceedings of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.6, T. Lambrou et al., "Classification of audio signals using statistical features on time and wavelet transform domains", pp.3621-3624, 12-15 May 1998, ISBN0-7803-4428-6, IEEE Catalog Number 98CH36181	1-47
A	Proceedings of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.2, H. Soltau et al., "Recognition of music types", pp.1137-1140, 12-15 May 1998, ISBN0-7803-4428-6, IEEE Catalog Number 98CH36181	1-47
A	Proceedings of the Fifth International Symposium on Signal Processing and Its Applications, ISSPA'99, Vol.1, K. Melih et al., "Audio source type segmentation using a perceptually based representation", pp.51-54, 22-25 August 1999, ISBN1-86435-451-8, IEEE Catalog Number 99EX359	1-47
A	Transactions of the Institute of Electronics, Information and Communication Engineers, Vol.J79-D-II, No.11,	1-47



Further documents are listed in the continuation of Box C.



See patent family annex.

\* Special categories of cited documents:  
 "A" document defining the general state of the art which is not considered to be of particular relevance  
 "E" earlier document but published on or after the international filing date  
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 "O" document referring to an oral disclosure, use, exhibition or other means  
 "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art  
 "&" document member of the same patent family

Date of the actual completion of the international search  
11 October, 2000 (11.10.00)Date of mailing of the international search report  
07 November, 2000 (07.11.00)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	November 1996, Kunio KASHIWANO et al., "Note recognition mechanisms in the OPTIMA processing architecture for music scene analysis", pp.1751-1761, issued on 25 November, 1996 (25.11.96), ISSN 0915-1923	
A	IPSJ Sig Notes [SIGMUS], Vol.97, No.18, 97-MUS-19-11, Kunio KASHIWANO et al., "Sound source identification using adaptive template mixtures - Formulation and application to music stream segregation -", pp.55-68, issued on 20 February, 1997 (20.02.97), ISSN 0919-6072	1-47
A	JP, 9-9199, A (Sony Corporation), 10 January, 1999 (10.01.99), Full text; all drawings (Family: none)	1-47
A	JP, 5-334861, A (Japan Radio Co., Ltd.), 17 December, 1993 (17.12.93), Full text; all drawings (Family: none)	1-47
A	JP, 7-105235, A (Sharp Corporation), 21 April, 1995 (21.04.95), Full text; all drawings & JP, 3021252, B2	1-47
A	JP, 8-265660, A (Nippon Telegr. & Teleph. Corp. <NTT>), 10 November, 1996 (10.11.96), Full text; all drawings (Family: none)	1-47
A	JP, 10-307580, A (Nippon Telegr. & Teleph. Corp. <NTT>), 17 November, 1998 (17.11.98), Full text; all drawings (Family: none)	1-47
A	JP, 10-319948, A (Nippon Telegr. & Teleph. Corp. <NTT>), 04 December, 1998 (04.12.98), Full text; all drawings (Family: none)	1-47

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int Cl' G10L15/00, G11B27/30, H04N5/76, H04N5/91

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int Cl' G10L15/00~17/00, G11B27/10~27/32,  
H04N5/76, H04N5/91

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926~1995年  
 日本国公開実用新案公報 1971~2000年  
 日本国登録実用新案公報 1994~2000年  
 日本国実用新案登録公報 1996~2000年

国際調査で使用了電子データベース (データベースの名称、調査に使用した用語)

JICST 科学技術文献ファイル (JOIS) INSPEC (DIALOG)  
 WPI (DIALOG)  
 IEEE/IEE Electronic Library

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	Proceedings of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.6, T.Lambrou et al, "Classification of audio signals using statistical features on time and wavelet transform domains", p.3621-3624, 12-15 May 1998, ISBN0-7803-4428-6, IEEE Catalog Number 98CH36181	1~47
A	Proceedings of 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, Vol.2, H.Soltan et al, "Recognition of music types", p.1137-1140, 12-15 May 1998, ISBN0-7803-4428-6, IEEE Catalog Number 98CH36181	1~47

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの  
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
 「O」 口頭による開示、使用、展示等に言及する文献  
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」 同一パテントファミリー文献

国際調査を完了した日

11.10.00

国際調査報告の発送日

07.11.00

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

松尾 淳 印

5C

8842

電話番号 03-3581-1101 内線 3540

## C (続き) . 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	Proceedings of the Fifth International Symposium on Signal Processing and Its Applications, ISSPA'99, Vol.1, K. Melih et al, "Audio source type segmentation using a perceptually based representation", p.51-54, 22-25 Aug. 1999, ISBN1-86435-451-8, IEEE Catalog Number 99EX359	1 ~ 4 7
A	電子情報通信学会論文誌 (Transactions of the Institute of Electronics, Information and Communication Engineers), Vol. J79-D-II, No.11, November 1996, 柏野邦夫外 (Kunio Kashiwano et al), 「音楽情景分析の処理モデルOPTIMAにおける単音の認識」 ("Note recognition mechanisms in the OPTIMA processing architecture for music scene analysis"), p.1751-1761, 1996年11月25日発行 (25.11.96), ISSN0915-1923	1 ~ 4 7
A	情報処理学会研究報告 [音楽情報科学] (IPSJ Sig Notes [SIGMUS]), Vol.97, No.18, 97-MUS-19-11, 柏野邦夫外 (Kunio Kashiwano et al), 「適応型混合テンプレートを用いた音源同定-複数楽器演奏への適用-」 ("Sound source identification using adaptive template mixtures - Formulation and application to music stream segregation -"), p.55-68, 1997年2月20日発行 (20.02.97), ISSN0919-6072	1 ~ 4 7
A	J P, 9-9199, A (ソニー株式会社) 10.1月.1999 (10.01.99) 全文全図 (ファミリーなし)	1 ~ 4 7
A	J P, 5-334861, A (日本無線株式会社) 17.12月.1993 (17.12.93) 全文全図 (ファミリーなし)	1 ~ 4 7
A	J P, 7-105235, A (シャープ株式会社) 21.4月.1995 (21.04.95) 全文全図 &JP, 3021252, B2	1 ~ 4 7
A	J P, 8-265660, A (日本電信電話株式会社) 10.11月.1996 (10.11.96) 全文全図 (ファミリーなし)	1 ~ 4 7
A	J P, 10-307580, A (日本電信電話株式会社) 17.11月.1998 (17.11.98) 全文全図 (ファミリーなし)	1 ~ 4 7
A	J P, 10-319948, A (日本電信電話株式会社) 4.12月.1998 (04.12.98) 全文全図 (ファミリーなし)	1 ~ 4 7